IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS:	**HARDIN ET AL.**	§	ART UNIT NO.: **1633**
SERIAL NO.:	**09/901,782**	§	EXAMINER: **SMITH, CL**
FILED:	**7/9/01**	§	DOCKET NO.: **00007/01UTL**
		§	
TITLE: **REAL-TIME SEQUENCE DETERMINATION**		§	
		§	

571-273-8300 FACSIMILE NO.

CERTIFICATE OF MAIL BY FACSIMILE TRANSMISSION

22 May 2007 Date of Deposit

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to the:

EXAMINER CL SMITH
MS AMENDMENT
571 273 8300

Robert W. Strozier

22 May 2007 Date of Signature

RULE 131 DECLARATION OF DR. SUSAN H. HARDIN

My name is Dr. Susan H. Hardin, and I am over 18 years of age and am a named inventor on this application. I am submitting this declaration to provide documentary support to antedate the cited Korlach et al. reference as it relates to beta or gamma labeled nucleotides.

I declare as follows:

I have thoroughly reviewed the cited Korlach et al. application, its parent application filed on 17 May 2000, and the 1999 provisional application from which it claims priority.

A thorough review of the Korlach et al. 1999 provisional application shows that the provisional application contains no disclosure of beta or gamma labeled nucleotides.

I have been advise by patent counsel that because the Korlach et al. 1999 provisional application does not disclose beta or gamma labeled nucleotides, documents dated prior to 17 May 2000, the filing date of the non-provisional Korlach et al. patent application, are all that is required to antedate the Korlach et al. reference as it relates to beta or gamma labeled nucleotides.

Attached is a document prepared and sent to a Federal Funding Agency prior to 17 May 2000 that disclosed the use of nucleotides labeled on the pyrophosphate, specifically the gamma or terminal phosphate, in sequencing strategies, including sequencing strategies based on measuring an interaction between a tag on the polymerase and a tag on the nucleotide, especially a tag on the pyrophosphate group - specifically the gamma phosphate.

Although the proposal is written in proposal format, the proposal lays out at least three inventions formulated in sufficient detail for an ordinary artisan to

understand.

1. Use of tagged polymerases and quencher tagged nucleotides in sequencing, where the quencher changes the fluorescence properties of the polymerase tag during an incorporation event
2. Use of tagged polymerases and tagged nucleotides in nucleotide sequencing, especially terminal or gamma phosphate tagged nucleotides, where the tags form a FRET pair.
3. Use of terminal or gamma phosphate tagged nucleotides in a sequencing based on the direct detection of released tagged pyrophosphate.

The funding agency's reviewers understood these three inventions and deemed them selectable. In addition, the inventors disclosed these and other inventions to VisiGen's patent counsel, Robert W. Strozier, prior to the 17 May 2000 date.

Upon instructions from patent counsel, the submitted document has been redacted to remove non relevant information and dates.

I hereby declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: **21 May 2007**

Respectfully submitted,

Susan H. Hardin

FedEx. *USA Airbill* FedEx Tracking Number

Form ID No.

Sender's Copy

1 From *Please print and press hard.*

Date

Sender's FedEx Account Number

Sender's Name Dr. S. Hardin Phone

Company UNIV OF HOUSTON

Address 3201 CULLEN BLVD

Dept./Floor/Suite/Room

City HOUSTON State TX ZIP 77004

2 Your Internal Billing Reference
First 24 characters will appear on invoice. OPTIONAL

3 To

Recipient's Name Phone

Company

Address

Dept./Floor/Suite/Room

City

A. TITLE: Real-time Sequence Determination
INSTITUTION: University of Houston, 4800 Calhoun, Houston, TX 77204-2163

PRINCIPAL INVESTIGATOR(S):

Principal Investigator: Susan H. Hardin
Address: Dept. of Biology/Biochem
University of Houston
Phone: 713-743-2686
Fax: 713-743-2636
Email: shardin@uh.edu
Date/Signature:

CO-Principal Investigator: James M. Briggs
Address: Dept. of Biology/Biochem
Address: University of Houston
Phone: 713-743-8366
Fax: 713-743-8351
Email: jbriggs@uh.edu
Date/Signature:

CO-Principal Investigator: Shiao-Chun Tu
Address: Dept. of Biology/Biochem
University of Houston
Phone: 713-743-8359
Fax: 713-743-8351
Email: dtu@uh.edu
Date/Signature:

CO-Principal Investigator: Xiaolian Gao
Address: Dept. of Chemistry
University of Houston
Phone: 713-743-2805
Fax: 240-250-5607
Email: xgao@uh.edu
Date/Signature:

CO-Principal Investigator: Richard Willson
Address: Dept. of Chemical Engineering
University of Houston
Phone: 713-743-4308
Fax: 713-743-4323
Email: willson@uh.edu
Date/Signature:

INSTITUTIONAL ENDORSEMENTS:

Administrative Head Betin Bilir-Santos
Address: Research Financial Services
4800 Calhoun, Houston,TX 77204
Phone: 713-743-3254
Fax: 713-743-9227
Email: bbilir@uh.edu
Date/Signature:

Institutional Financial Officer Betin Bilir-Santos
Address: Research Financial Services
4800 Calhoun, Houston,TX 77204
Phone: 713-743-3254
Fax: 713-743-9227
Email: bbilir@uh.edu
Date/Signature:

TOTAL COST OF PROJECT:

TOTAL GOVERNMENT FUNDING REQUESTED:
Dollar amount requested (sum of direct plus indirect costs)

PROPOSED START DATE:

PROPOSED END DATE:

DATE SUBMITTED:

THIS PROPOSAL IS SUBMITTED PURSUANT TO THE RESEARCH ANNOUNCEMENT

B. Table of Contents

C. PROJECT OBJECTIVES:

We propose to determine whether an individual DNA molecule can be directly used to rapidly produce accurate sequence information. To address this question we will develop a method that enables real-time single-molecule DNA sequencing. In this method a single tag that is strategically positioned on a DNA polymerase interacts with a color-coded dNTP. As the correct dNTP is incorporated during the polymerization reaction, the identity of the base is indicated by a signature fluorescent signal. The rate of polymerase incorporation can be varied, but is controlled to create a 'real-time' readout of polymerase activity and base sequence. Sequence data can be collected at a rate of >100,000 bases per hour from each reaction.

Development of single-molecule DNA sequencing requires, as a first step, identifying the optimal DNA polymerase for genetic engineering. Subsequently, candidate amino acids within the polymerase that can be mutated and modified without significantly affecting polymerization efficiency will be identified (computational analysis, Dr. J. Briggs). These identified amino acids will be genetically engineered to facilitate dye attachment. However, since each modified site may differentially affect activity, several candidate sites will be individually altered. These enzymes will be expressed, purified, and assayed for activity (molecular biology, Dr. S. Hardin). The chemical attachment of a fluorescence donor to the engineered polymerase and characterization of the modified enzymes will be carried out in Dr. Tu's lab. Once the optimized enzyme is identified, it will be used to stimulate fluorescence transfer with an incoming dNTP (design of detection equipment, Dr. R. Willson; choices of fluorescent donors and acceptors, Drs. D. Tu and X. Gao; choice of site for labeling dNTP, Drs. Briggs, Hardin, Tu, Gao). These assays will enable us to determine the identity of incorporated (tagged) dNTPs. Simple sequences will be used as templates in our initial studies, and more complex templates will be introduced as the project reaches designated milestones. Concurrently, we will identify tagged dNTPs that work optimally in our real-time DNA sequencing system (synthesis of tagged dNTPs, Dr. X. Gao) and develop software that will analyze the fluorescence emitted from the reaction and interpret base identity (Drs. J. Briggs and S. Hardin).

D. Statement of the Approach

We have assembled a research team with complementary areas of expertise in 1) Molecular Biology, Biochemistry, and Chemistry; 2) Computer Science; and 3) Chemical and Mechanical Engineering:

Dr. Susan Hardin provides expertise in molecular biology, DNA replication, and DNA sequencing. Dr. Hardin's group will identify the optimal enzyme to use for our studies. They will also genetically modify the gene encoding DNA polymerase, sequence the resulting polymerase clones, and assay enzyme activity.

Dr. Shiao-Chun (David) Tu provides expertise in energy transfer reactions, as well as protein purification and enzymology. Dr. Tu's group will identify optimal dyes for both enzyme and dNTP fluorescent-tagging experiments. They will also be responsible for fluorescently modifying, purifying, and characterizing engineered polymerase.

Dr. Xiaolian Gao provides expertise in chemical synthesis of unusual deoxynucleoside triphosphates. Dr. Gao's group will design, synthesize, and purify tagged dNTPs (base, sugar, or phosphate labeled).

Dr. James Briggs provides computational expertise. Dr. Briggs' effort includes identification of candidate amino acids for targeted mutagenesis of the polymerase via modeling of the complex between the (labeled) dNTP and the (labeled) protein. The efficiency of the fluorescence resonance energy transfer (FRET) will be predicted. Dr. Briggs' group will also work closely with Dr. Hardin's group to create the base identification software.

Dr. Richard Willson provides expertise in fluorescence, as well as chemical- and instrument-engineering. Dr. Willson's group will be responsible for optimizing larger-scale expression and purification of the polymerase. They will also identify and develop equipment that will meet our needs for both development and single-molecule detection stages of the project.

PROJECT MILESTONES TOWARD REAL-TIME SEQUENCE DETERMINATION:

- Identify the most appropriate polymerase for our studies (Hardin).
- Structural modeling and interpretation of results identify the optimal modification site for tag attachment on the polymerase (Briggs, Hardin).
- Engineer, express, and purify the polymerase (Hardin).
- Fermentation, purification, and quality control of polymerase candidate protein (Willson).
- Identify and/or design optimal fluorescence dyes for the attachment to the engineered polymerase (Gao, Tu).
- Biochemical, enzymological, and sequencing performance characterization of novel polymerases (Willson, Tu, Hardin).
- Molecular modeling will provide a rational method for identifying the best sites for fluorescently labeling the dNTP, and for estimating fluorescence resonance energy transfer (FRET) efficiency (Briggs, Gao, Tu).
- Identify and/or design optimal labeling system (Gao, Tu).
- Characterization of FRET behavior of labeled polymerase/labeled substrate systems (Tu, Willson)
- Design detection systems. One will be used to assay progress in enzyme design and modification using non-challenging detection methods. The second will be a prototype that will unite technologies and enable single molecule detection. Construction and optimization of single-molecule fluorescence apparatus and techniques (Willson, Tu)
- Develop a computer algorithm to interpret the fluorescent signals and present the user with DNA sequence information (Briggs, Hardin).
- Overall coordination of project efforts (Hardin) active communication and optimization of interdependencies of the project (Hardin, Briggs, Gao, Tu, Willson).

E. SIGNIFICANCE:

Engineering a polymerase to function as a direct molecular sensor of DNA base identity allows us to create the fastest enzymatic DNA sequencing system possible. At this point, direct readout from a polymerase to determine base sequence is a 'virtual' invention. Several variations of a basic method are envisioned. Development of this method will impact other disciplines. The proposed method will enable new ways to address basic research questions that extend beyond monitoring conformational changes occurring during replication or assaying polymerase incorporation fidelity in a variety of sequence contexts. The technologies developed during the course of this work will facilitate single-molecule detection systems, fluorescent molecule chemistry, computer modeling and base-calling algorithms, and genetic engineering of biomolecules. If we are successful, these methods will be invaluable. They have the potential of replacing current DNA sequencing technologies. They may make it easier to classify an organism or identify variations within an organism by simply sequencing

[redacted]

ADVANTAGES OF REAL-TIME SEQUENCE DETERMINATION:

- This strategy eliminates sequencing reaction processing, gel or capillary loading, electrophoresis, and data assembly, promising huge savings in labor, time, and cost.
- Real-time data determination.
- Ability to process many samples in parallel.
- Sequence a genome in a day or less [redacted] and characterization).
- Greater than 2 orders of magnitude increase in sequence throughput anticipated *per reaction*.
- Diagnostic uses, i.e. Single Nucleotide Polymorphism (SNP) detection.
- Basic research applications (i.e. examination of polymerase incorporation rates in a variety of different sequence contexts; analysis of errors in different contexts; epigenotypic analysis).
- Enabling technology for creation of a robust (rugged) single molecule detection system.
- Development of systems and procedures that are compatible with biomolecules.
- Pushes development of genetic nanotechnology.

F. THE IDEA

A brief overview of the proposed single-molecule DNA sequencing process follows: We envision placing a single tag on the polymerase and a unique tag on each dNTP. As a tagged dNTP is incorporated into the DNA polymer, a characteristic fluorescent signal is emitted that indicates base identity (emission wavelength and/or strength provide signature for base identity). Tagged dNTPs will be identified that do not interfere with Watson-Crick base pairing or significantly impact polymerase incorporation. Initially, we will focus on dyes used to fluorescently label ddNTPs for automated DNA sequencing, since they are incorporated in a template-directed manner by the polymerase. Additionally, we will determine whether dNTPs containing tags attached to the terminal (gamma) phosphate are directly detected upon incorporation (four color, base-specific phosphate cleavage stimulates detector). An advantage of this latter approach is that the nascent DNA strand will not contain fluorescent bases and, therefore, should produce minimal enzyme distortion and background fluorescence. The fluorescent signals produced upon incorporation will be detected and analyzed to determine DNA base sequence.

INTRODUCTION AND BACKGROUND INFORMATION

Overview of Conventional DNA Sequencing

The development of methods that allow one to quickly and reliably determine the order of bases or 'sequence' in a fragment of DNA is a key technical advance, the importance of which cannot be overstated. Knowledge of DNA sequence enables a greater understanding of the molecular basis of life. DNA sequence information provides scientists with information critical to a wide range of biological processes. The order of bases in DNA specifies the order of bases in RNA, the molecule within the cell that directly encodes the informational content of proteins. DNA sequence information is routinely used to deduce protein sequence information. Base order dictates DNA structure and its function, and provides a molecular program that can specify normal development, manifestation of a genetic disease, or cancer.

Knowledge of DNA sequence and the ability to manipulate these sequences has accelerated development of biotechnology and led to the development of molecular techniques that provide the tools to ask and answer important scientific questions. The polymerase chain reaction (PCR), an important biotechnique that facilitates sequence-specific detection of nucleic acid, relies on sequence information. DNA sequencing methods allow scientists to determine whether a change has been introduced into the DNA, and to assay the effect of the change on the biology of the organism, regardless of the type of organism that is being studied. Ultimately, DNA sequence information may provide a way to uniquely identify individuals.

In order to understand the DNA sequencing process, one must recall several facts about DNA. First, a DNA molecule is comprised of four bases, adenine (A), guanine (G), cytosine (C), and thymine (T). These bases interact with each other in very specific ways through hydrogen bonds, such that A interacts with T, and G interacts with C. These specific interactions between the bases are referred to as base-pairings. In fact, it is these base-pairings (and base stacking interactions) that stabilize double-stranded DNA. The two strands of a DNA molecule occur in an antiparallel orientation, where one strand is positioned in the 5' to 3' direction, and the other strand is positioned in the 3' to 5' direction. The terms 5' and 3' refer to the directionality of the DNA backbone, and are critical to describing the order of the bases. The convention for describing base order in a DNA sequence uses the 5' to 3' direction, and is written from left to right. Thus, if one knows the sequence of one DNA strand, the complementary sequence can be deduced.

Sanger DNA Sequencing (Enzymatic Synthesis)

Sanger sequencing is currently the most commonly used method to sequence DNA (Sanger *et al.*, 1977). This method exploits several features of a DNA polymerase: its ability to make an exact copy of a DNA molecule, its directionality of synthesis (5' to 3'), its requirement of a DNA strand (a 'primer') from which to begin synthesis, and its requirement for a 3' OH at the end of the primer. If a 3' OH is not available, then the DNA strand cannot be extended by the polymerase. If a dideoxynucleotide (ddNTP; ddATP, ddTTP, ddGTP, ddCTP), a base analogue lacking a 3' OH, is added into an enzymatic sequencing reaction, it is incorporated into the growing strand by the polymerase. However, once the ddNTP is incorporated, the polymerase is unable to add any additional bases to the end of the strand. Importantly, ddNTPs are incorporated by the polymerase into the DNA strand using the same base incorporation rules that dictate incorporation of natural nucleotides, where A specifies incorporation of T, and G specifies incorporation of C (and vice versa).

Fluorescent DNA Sequencing

A major advance in determining DNA sequence information occurred with the introduction of automated DNA sequencing machines (Smith *et al.*, 1986). The automated sequencer is used to separate sequencing reaction products, detect and collect (via computer) the data from the reactions, and analyze the order of the bases to automatically deduce the base sequence of a DNA fragment. Automated sequencers detect extension products containing a fluorescent tag. Sequence read lengths obtained using an automated sequencer are dependent upon a variety of parameters, but typically range between 500 to 1,000 bases (3-18 hours of data collection). At maximum capacity an automated sequencer can collect data from 96 samples in parallel.

When dye-labeled terminator chemistry is used to detect the sequencing products, base identity is determined by the color of the fluorescent tag attached to the ddNTP. After the reaction is assembled and processed through the appropriate number of cycles (3-12 hours), the extension products are prepared for loading into a single lane on an automated sequencer (unincorporated, dye-labelled ddNTPs are removed and the reaction is concentrated; 1-2 hours). An advantage of dye-terminator chemistry is that extension products are visualized only if they terminate with a dye-labelled ddNTP; prematurely terminated products are not detected. Thus, reduced background noise typically results with this chemistry.

State-of-the-art dye-terminator chemistry uses four energy transfer fluorescent dyes (Rosenblum *et al.*, 1997). These terminators include a fluorescein donor dye (6-FAM) linked to one of four different dichlororhodamine (dRhodamine) acceptor dyes. The dRhodamine acceptor dyes associated with the terminators are dichloro[R110], dichloro[R6G], dichloro[TAMRA] or dichloro[ROX], for the G-, A-, T- or C-terminators, respectively. The donor dye (6-FAM) efficiently absorbs energy from the argon ion laser in the automated sequencing machine and transfers that energy to the linked acceptor dye. The linker connecting the donor and acceptor portions of the terminator is optimally spaced to achieve essentially 100% efficient energy transfer. The fluorescence signals emitted from these acceptor dyes exhibit minimal spectral overlap and are collected by an ABI PRISM 377 DNA sequencer using 10 nm virtual filters centered at 540, 570, 595 and 625 nm, for G-, A-, T- or C-terminators, respectively. Thus, energy transfer dye-labeled terminators produce brighter signals and improve spectral resolution. These improvements result in more accurate DNA sequence information.

The predominant enzyme used in automated DNA sequencing reactions is a genetically engineered form of DNA polymerase I from *Thermus aquaticus*. This enzyme, Ampli*Taq* DNA Polymerase, FS, was optimized to more efficiently incorporate ddNTPs and to eliminate the 3' to 5' and 5' to 3' exonuclease activities. Replacing a naturally occurring phenylalanine at position 667 in *T. aquaticus* DNA polymerase with a tyrosine reduced the preferential incorporation of a dNTP, relative to a ddNTP (Tabor and Richardson, 1995; Reeve and Fuller, 1995). Thus, a single hydroxyl group within the polymerase is responsible for discrimination between dNTPs and ddNTPs. The 3' to 5' exonuclease activity, which enables the polymerase to remove a mis-incorporated base from the newly replicated DNA strand (proofreading activity), was eliminated because it also allows the polymerase to remove an incorporated ddNTP. The 5' to 3' exonuclease activity was eliminated because it removes bases from the 5' end of the reaction products. Since the reaction products are size separated during gel electrophoresis, interpretable sequence data is only obtained if the reaction products share a common endpoint. More specifically, the primer defines the 5' end of the extension product and the incorporated, color-coded ddNTP defines base identity at the 3' end of the molecule. Thus, conventional DNA sequencing involves analysis of a population of DNA molecules sharing the same 5' endpoint, but differing in the location of the ddNTP at the 3' end of the DNA chain.

Genome Sequencing

Very often a researcher needs to determine the sequence of a DNA fragment that is larger than the 500-1,000 base average sequencing read length. Not surprisingly, strategies to accomplish this have been developed. These strategies are divided into two major classes, random or directed, and strategy choice is influenced by the size of the fragment to be sequenced.

In random or shotgun DNA sequencing, a large DNA fragment (typically one larger than 20,000 base pairs) is broken into smaller fragments that are inserted into a cloning vector. It is assumed that the sum of information contained within these smaller clones is equivalent to that contained within the original DNA fragment. Numerous smaller clones are randomly selected, DNA templates are prepared for sequencing reactions, and primers that will base-pair with the vector DNA sequence bordering the insert are used to begin the sequencing reaction (2-7 days for a 20 kbp insert). Subsequently, the quality of each base call is examined (manually or automatically *via* software (PHRED, Ewing *et al.*, 1998); 1-10 minutes per sequence reaction), and the sequence of the original DNA fragment is reconstructed by computer assembly of the sequences obtained from the smaller DNA fragments. Based on the time estimates provided, if a shotgun sequencing strategy is used, a 20 kbp insert is expected to be completed in 3-10 days. This strategy is being extensively used to determine the sequence of ordered fragments that represent the entire human genome (http://www.nhgri.nih.gov/HGP/). However, this random approach is typically not sufficient to complete sequence determination, since gaps in the sequence often remain after computer assembly. A directed strategy (described below) is usually used to complete the sequence project.

A directed or primer-walking sequencing strategy can be used to fill-in gaps remaining after the random phase of large-fragment sequencing, and as an efficient approach for sequencing smaller DNA fragments. This strategy uses DNA primers that anneal to the template at a single site and act as a start site for chain elongation. This approach requires knowledge of some sequence information to design the primer. The sequence obtained from the first reaction is used to design the primer for the next reaction and these steps are repeated until the complete sequence is determined. Thus, a primer-based strategy involves repeated sequencing steps from known into unknown DNA regions, the

process minimizes redundancy, and it does not require additional cloning steps. However, this strategy requires the synthesis of a new primer for each round of sequencing.

The necessity of designing and synthesizing new primers, coupled with the expense and the time required for their synthesis, has limited the routine application of primer-walking for sequencing large DNA fragments. Researchers have proposed using a library of short primers to eliminate the requirement for custom primer synthesis (Studier, 1989; Siemieniak and Slightom, 1990; Kieleczawa *et al.*, 1992; Kotler *et al.*, 1993; Burbelo and Iadarola, 1994; Hardin *et al.*, 1996; Raja *et al.*, 1997; [redacted]). The availability of a primer library minimizes primer waste, since each primer is used to prime multiple reactions, and allows immediate access to the next sequencing primer.

One of the original goals of the Human Genome Project was to complete sequence determination of the entire human genome by 2005 (http://www.nhgri.nih.gov/HGP/). However, the plan is ahead of schedule and a 'working draft' of the human genome will be completed by 2001 (Collins *et al.*, 1998). Due to technological advances in several disciplines, the completed genome sequence is expected in 2003, two years ahead of schedule. Progress in all aspects involving DNA manipulation (especially manipulation and propagation of large DNA fragments), evolution of faster and better DNA sequencing methods (http://www.abrf.org), development of computer hardware and software capable of manipulating and analyzing the data (bioinformatics), and automation of procedures associated with generating and analyzing DNA sequences (engineering) are responsible for this accelerated time frame.

Single-Molecule DNA Sequencing

Conventional DNA sequencing strategies and methods are reliable, but time, labor, and cost intensive. To begin to address these issues, some researchers are investigating fluorescence-based, single-molecule sequencing methods utilizing enzymatic degradation, followed by single-dNMP detection and identification (Davis *et al.*, 1991; Davis *et al.*, 1992; Keller *et al.*, 1996; Goodwin *et al.*, 1997; [redacted]). However, we believe that by engineering the polymerase to function as a *direct* molecular sensor of DNA base identity, we will be able to create the fastest and most efficient enzymatic DNA sequencing system possible. At this point, direct readout from a polymerase to determine base sequence is a 'virtual' invention, but, once developed, it will enable new ways to address basic research questions that extend beyond monitoring conformation changes occurring during replication or assaying polymerase incorporation fidelity in a variety of sequence contexts. Development of these methods will impact other disciplines. The technologies developed and optimized during the course of this project will facilitate single-molecule detection systems, fluorescent molecule chemistry, computer modeling and base-calling algorithms, and genetic engineering of biomolecules. If we are successful, these methods will be invaluable.

Single-molecule DNA sequencing has the potential to replace current DNA sequencing technologies. It is projected to decrease time, labor, and costs associated with the sequencing process. The technology also promises to be highly scalable. Single-molecule DNA sequencing has the potential to increase the DNA sequence discovery process by at least two orders of magnitude per reaction. Single-molecule DNA sequencing may make it easier to classify an organism or identify variations within an organism by simply sequencing the genome in question. This application may be

[redacted]

Our Approach

A brief overview of the proposed single-molecule DNA sequencing process follows: In the first approach, we envision placing a fluorescence donor on the polymerase (i.e. fluorescein or fluorescein-type molecule) and a fluorescence acceptor with a unique fluorescent tag color on each dNTP (i.e. d-rhodamine or similar molecule). As incoming fluorescently-tagged dNTP is bound by the polymerase for DNA elongation, a characteristic fluorescent signal is emitted that indicates base identity (emission wavelength and intensity provide signature for base identity). Fluorescently-tagged dNTPs will be identified that do not interfere with Watson-Crick base pairing or significantly impact polymerase incorporation. Initially, we will focus on dyes used to fluorescently label ddNTPs for automated DNA sequencing, since they are incorporated in a template-directed manner by the polymerase. Additionally, we will determine whether dNTPs containing fluorescent tags attached to the terminal (gamma) phosphate are directly detected upon incorporation (four color, base-specific phosphate cleavage stimulates detector). An advantage of this latter approach is that the nascent DNA strand will not contain fluorescent bases and, therefore, should produce minimal enzyme distortion and background fluorescence. A second approach will be using fluorescently labeled polymerase as before. However, the dNTPs will be labeled with different quenchers for the fluorescence tag on the polymerase. Each of these quenchers should have distinguishable degrees of quenching efficiencies. Consequently, the identity of each incoming labeled dNTP can be determined by its unique efficiency in quenching the emission of the fluorescently labeled polymerase. The signals produced during incorporation will be detected and analyzed to determine DNA base sequence.

TECHNICAL APPROACH AND CONSIDERATIONS

Enzyme Choice

Our choice of polymerase for the development of single-molecule DNA sequencing methods is critical. All subsequent work depends on this choice. Thus, the reasons we chose to genetically engineer the DNA polymerase from *Thermus aquaticus* - *Taq* DNA polymerase - for our studies are listed and, subsequently, discussed in more detail.

- **Crystal structures are available for this enzyme**
- **Efficiently expressed in *E. coli***
- **No cysteines are present in the protein sequence**
- **The processivity of the enzyme can be modified**
- **This polymerase lacks a 3' to 5' exonuclease activity**
- **It possesses a 5' to 3' exonuclease activity**
- ***Taq* DNA polymerase is thermostable**
- **Error rates are characterized**

Crystal structures are available for Taq DNA polymerase

The enzyme chosen for single-molecule sequencing development should be one for which a crystal structure is solved. Knowledge of protein structure enables a more informed choice of candidate amino acids within the polymerase to alter for fluorescent tag attachment without adversely affecting polymerase activity. There are 14 structures solved for *Taq* DNA polymerase, either with or without DNA template/primer, dNTP, or ddNTP, making this enzyme an excellent candidate for our studies (Eom *et al.*, 1996;Li *et al.*, 1998; Li *et al.*, 1998).

Taq DNA polymerase is efficiently expressed in E. coli

The protein used to form the crystals for these structures was produced by over-expression in *E. coli*. Although an apparently minor detail, it is essential that we are able to efficiently produce and purify many variants of the enzyme to more rapidly identify and characterize the optimal enzyme for single-molecule DNA sequencing.

No cysteines are present in the protein sequence

An additional advantage of working with *Taq* DNA polymerase is that its protein sequence lacks cysteines. Thus, our choices of amino acids to target for mutation are not limited and subsequent modification is simplified. As is discussed below, cysteine is the site at which the enzyme will be fluorescently labeled.

The processivity of the enzyme can be modified

Although crystal structures are available and the enzyme does not contain naturally occurring cysteines, native *Taq* DNA polymerase is not optimally suited for our purposes since it is not a very processive polymerase (50-80 nucleotides are incorporated before dissociation). It can, however, be appropriately engineered. Specifically, development of a single-molecule DNA sequencer will benefit by using a DNA polymerase that remains associated with the DNA template during the extension phase of the sequencing reaction. Using a highly processive enzyme is expected to minimize complications that may arise from dissociation from the template, which will alter the polymerization rate. However, these rate differences could be compensated for by appropriately modifying the base calling software. Thus, lack of processivity may not limit the sequence lengths achievable by this invention.

This feature - processivity - of the native *Taq* enzyme could negatively impact sequencing run lengths. However, enzymes responsible for replicating the genome are very processive and are able to replicate thousands of bases before dissociating from the template (Kornberg and Baker, 1992). In fact, eukaryotic and prokaryotic DNA polymerases possess mechanisms to overcoming this shortcoming: Increased processivity is achieved through the use of accessory factors (Kelman *et al.*, 1998). A particularly relevant example involves T7 DNA polymerase and its interaction with thioredoxin, a 12 kDa protein produced by *E. coli*. These proteins associate to form a complex that effectively encircles the DNA template, anchoring the replication complex to the template and achieving a several thousand-fold increase in processivity of T7 DNA polymerase (Tabor *et al.*, 1987; Huber *et al.*, 1987).

Processivity can also be altered through genetic engineering, as was elegantly demonstrated using the Klenow fragment from *E. coli* DNA polymerase I, a polymerase with even lower processivity than *Taq*. Increased processivity was obtained by introducing the 76 amino acid 'processivity domain' from T7 DNA polymerase into the Klenow fragment (Bedford *et al.*, 1997; [redacted]. More specifically, this processivity domain contains the thioredoxin binding domain (TBD) from T7 DNA polymerase and it was engineered into the Klenow fragment between the H and H_1 helices (at the tip of 'thumb' region within the polymerase). This sequence addition caused a thioredoxin-dependent increase in both the processivity and specific activity of Klenow fragment. Thus, we propose to introduce this same region of T7 DNA polymerase into the homologous site of *Taq* DNA polymerase [redacted]. If necessary, the TBD and thioredoxin can be altered to become more heat stable.

Taq DNA polymerase possesses a 5' to 3' exonuclease activity and is thermostable

A DNA polymerase must have access to the replication template – a single-stranded DNA molecule - to polymerize a nascent DNA strand. In conventional DNA sequencing, creating single-stranded molecules is accomplished by heat denaturing complementary DNA strands (in cycle sequencing reactions). We propose to begin our studies by using single-stranded M13 DNA as sequencing templates and appropriate synthetic oligonucleotide primers. Initially, template regions will require extension through simple sequences (defined lengths of either homo-polynucleotide or di-nucleotide repeats in templates) and sites of initiation will be defined by synthetic oligonucleotides. Ultimately, however, we must develop methods that allow us to directly determine sequence information from an isolated chromosome – a double-stranded DNA molecule. We envision that heating this sample may not be sufficient to produce or maintain a single-stranded DNA molecule.

To favor the single-stranded state, we propose retaining the 5' to 3' exonuclease activity of the native *Taq* DNA polymerase in the enzyme engineered for single-molecule DNA sequencing. This activity cleaves 5' terminal nucleotides from double-stranded DNA (*via* phosphodiester bond hydrolysis) and releases mono- and oligonucleotides (Holland *et al.*, 1991). It is this activity of the polymerase that is exploited by the '*Taq*Man' assay. The presence of this exonuclease activity will enable the polymerase to remove a duplex strand that may renature downstream from the replication site using a nick-translation reaction mechanism. If fact, it will be interesting to determine whether the single-molecule sequencing method will require a synthetic oligonucleotide primer to initiate the reaction, or whether a nick in the DNA molecule can serve as the site for reaction initiation.

The polymerase lacks a 3' to 5' exonuclease activity

Taq DNA polymerase lacks a 3' to 5' exonuclease activity (proofreading activity). This is important for our studies since we do not want the enzyme to remove a base for which fluorescent signal was detected. If the enzyme used in single-molecule DNA sequencing possessed a 3' to 5' exonuclease activity, the enzyme would add another base to replace of the one that had been removed. This newly added base would produce a signature fluorescent signal that would suggest the presence of two identical bases in the template. This type of artifact could be detrimental to the technology.

All polymerases make replication errors. The 3' to 5' exonuclease activity is used to proofread the newly replicated DNA strand. Since *Taq* DNA polymerase lacks this proofreading function, an error in base incorporation becomes an error in DNA replication. Error rates for *Taq* DNA polymerase are 1 error per ~100,000 bases synthesized (Eckert and Kunkel, 1990; Cline *et al.*, 1996). The enzyme achieves relatively high fidelity synthesis by inefficiently incorporating non-complementary dNTPs and/or poorly extending a mismatched primer/template (Cline *et al.*, 1996). *Pfu* DNA polymerase is a thermostable polymerase possessing a 3' to 5' proofreading activity (Lundberg *et al.*, 1991). Data from a *Pfu* DNA polymerase variant lacking proofreading activity suggests that this activity reduces the ability of polymerase to discriminate between bases (Cline *et al.*, 1996). These researchers also determined that reaction conditions affect the error rate of the enzyme, demonstrating the importance of reaction optimization on sequence accuracy.

Thus, polymerase error rate will not negatively impact or limit the length of sequence attainable by the single-molecule DNA sequencing system. However, we will determine the error rate of our system through comparisons with known sequences. This information is essential for

determining whether reactions should be processed in parallel and, if so, the optimal number that should be processed to assign confidence values to the sequence data. For example, we may observe that base context influences polymerase accuracy and this information may enable us to assign confidence values to individual base calls. However, depending on the goal of a particular sequencing project, it may be more important to generate a genome sequence as rapidly as possible. For example,

[redacted]

Taq DNA polymerase is the enzyme of choice for single-molecule DNA sequencing

Engineering the polymerase to function as a direct molecular sensor of DNA base identity allows us to create the fastest enzymatic DNA sequencing system possible. For the reasons detailed above, *Taq* DNA polymerase is the optimal enzyme to genetically modify and adapt for single-molecule DNA sequencing. Additionally, basic research questions concerning DNA polymerase structure and function during replication can be addressed using this technology. Advances in single molecule detection technology and molecular modeling will enable similar advances in other disciplines.

SELECTION OF SITES FOR PROTEIN MUTATION TO ACCEPT FLUORESCENT TAG

A key task for this project is the identification of amino acids in the polymerase that can withstand mutation and fluorescent labeling. This will be accomplished via a combination of computational methods, mutational studies, and assaying for normal protein function. Follow-up computational analyses will be performed to refine the molecular models such that they might be used to help suggest alternative sites for incorporation of a fluorescent tag in the event that problems are encountered with the preliminary suggestions.

The identification of sites in the polymerase that are not in contact with other proteins, that should not alter the conformation or folding of the protein, and that are not involved in the function of the protein, will be accomplished by a combination of sequence analyses and molecular docking studies. Regions of the protein surface that are not important for function can be identified, indirectly, by investigating the variation in sequence as a function of evolutionary time and protein function, with use of the *evolutionary trace* method (Lichtarge *et al.*, 1996). In this approach, amino acid residues that are important for structure or function are found by comparing evolutionary mutations and structural homologies. The polymerases are ideal systems for this type of study, as there are many crystal and co-crystal structures and many available sequences. We will exclude the regions of structural/functional importance from consideration as sites for mutation/labeling. In addition, visual inspection and overlays of available structures in different conformational states, as already available from crystallographic studies, will further assist in identifying areas near the binding site for dNTPs that might be available for mutation and labeling. We envision choosing amino acids somewhat internally located, perhaps surrounding the enzyme active site, to reduce background (i.e. enzyme interacting with non-specifically associated dNTPs). Mutated and labeled polymerases will be built and energy minimized in a full solvent environment to estimate the effect on the structure of the mutation and/or labeling. This will also provide an estimate of the orientation of the fluorescent label with respect to the dNTP-binding pocket, thereby allowing us to estimate the FRET efficiency prior to measurement.

One of the major difficulties that will be encountered with the modeling studies will be due to the lack of molecular mechanics force field parameters for the fluorescent tags and for the fluorescently tagged amino acid (i.e. protein) or dNTP. Force field parameters must therefore be developed for these studies. The parameter development will be accomplished in the usual way (MacKerrell *et al.*, 1998) by employing a combination of quantum mechanical studies to obtain partial charge distributions and energies for relevant intramolecular conformations (i.e. for the dihedral angle definitions). Dr. Briggs has many years of experience in force field parameterization, starting with his doctoral studies on the generation of molecular mechanics parameters for small organic molecules. [redacted] molecular modeling computer programs will be used to generate initial topologies and structures for the new molecules.

The molecular models generated in the early phases of this project will be continually refined to take into account new experimental data and then will be used to make subsequent predictions to support continuing phases of the project. This refinement will involve more sophisticated molecular refinement methods (e.g. molecular dynamics) and adjustments in force field parameterization based on continued testing.

Preliminary results

The Cartesian coordinates for GTP were obtained and used in a manual docking experiment to generate a complex between GTP and a DNA polymerase I. The X-ray structure for the polymerase was that from *B. stearothermophilus* with a DNA primer template bound (Kiefer *et al.*, 1998; pdb code: 2bdp). The GTP was manually placed in the proposed dNTP binding site according to the procedure described in Kiefer *et al.*, 1998 (see Figure 4 and associated description in the text). The most relevant points are that at least one oxygen from each phosphate in GTP was within ca. 3.0Å of the observed Mg^{2+} ion and that the base partially stacks with the base at the end of the primer strand.

Figure: DNA polymerase I from *B. stearothermophilus* (Kiefer *et al.*, 1998; pdb code: 2bdp) co-crystallized with DNA template primer and with a manually docked GTP molecule. The protein is represented by a blue solid tube and the DNA by a yellow backbone trace and bonds. The GTP is near the center of each image represented in ball and stick. The rust sphere near the phosphates in GTP is the Mg^{2+} ion bound to the highly conserved Asp653 and Asp830 protein sidechains. In the right most image, the molecular surface of the protein is displayed along with DNA and GTP. GTP is atom colored and Mg^{2+} is rendered in black.

It is clear from the initial molecular modeling studies that the dNTP can be labeled on sites other than the traditional 7-position on purines and the 5-position on pyrimidines. One of the ideas presented in this proposal is to put the fluorescent tag on the γ-phosphate such that, upon base incorporation, the tagged PPi will diffuse away from the protein (i.e. FRET will cease). According to our preliminary

modeling studies, and the GTP/protein complex model presented in Figure 4 of Kiefer *et al.*, 1998, there appears to be sufficient room for a tag on the γ-phosphate, without inhibiting incorporation.

Selection of site in dNTP to accept fluorescent tag

Molecular docking simulations will be carried out to predict the docked orientation of the natural and fluorescently labeled dNTPs using the AutoDock computer program (Morris *et al.*, 1998; [redacted]. Conformational flexibility will be permitted during the docking simulations making use of an efficient Lamarckian Genetic algorithm implemented in the AutoDock program. A subset of protein sidechains can also be allowed to move to accommodate the dNTP as it docks. The best docked configurations will be energy minimized in the presence of a solvent environment. Experimental data are available which identify amino acids in the polymerase active site that are involved in catalysis and in contact with the template/primer DNA strands or the dNTP to be incorporated. The docking studies will help us to support which sites in the dNTP can be labeled and to predict the FRET efficiency that we might expect.

Prediction of FRET efficiency

The efficiency of the FRET that we might expect to see will be estimated according to the double helical model previously proposed (Furey *et al.*, 1998; Clegg *et al.*; 1993). The efficiency of energy transfer (E) can be computed from the following equation:

$$E = 1/(1+[R/R_0]^6)$$

where R_0 is the Förster critical distance at E=0.5 and is calculated from:

$$R_0 = (9.79x10^3)(\kappa^2 n^{-4} Q_D J_{DA})^{1/6}$$

n is the refractive index of the medium (1.4 for aqueous solution), κ^2 is a geometric orientation factor related to the relative angle of the two transition dipoles (assumed to be 2/3 here), J_{DA} ($M^{-1}cm^3$) is the overlap integral representing the normalized spectral overlap of the donor emission and acceptor absorption, and Q_D is the quantum yield. The overlap integral can be computed from:

$$J_{DA} = [\int F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda]/[\int F_D(\lambda)d\lambda]$$

where F_D is the donor emission, ε_A is the acceptor absorption. Q_D is obtained in the following way:

$$Q_D = Q_{RF}(I_D/I_{RF})(A_{RF}/A_D)$$

where I_D and I_{RF} are the fluorescence intensities of donor and a reference compound (fluorescein in 0.1N NaOH), and A_{RF} and A_D are the absorbances of the reference compound and donor. Q_{RF} is the quantum yield of fluorescein in 0.1N NaOH and is taken to be 0.90.

R, the distance between the donor and acceptor, can then be measured by looking at different configurations (e.g. conformations) of the labeled protein and labeled dNTP, in order to obtain a conformationally averaged value. The R_0 will be estimated from the equations above, although was found to be 58.5 Å in the previous study with the donor attached to Cys751 of the protein (Furey *et al.*, 1998). We will also be in a position to estimate the appropriateness of the alignment of the transition dipoles, which is a key factor affecting the transfer efficiency.

MUTAGENESIS AND SEQUENCING OF POLYMERASE VARIANTS

The gene encoding *Taq* DNA polymerase was obtained from [redacted] and will be expressed in [redacted] *E. coli* strain DH1 (Engelke *et al.*, 1990). Once candidate amino acids are identified for mutagenesis, we will use standard molecular methods to introduce a cysteine codon, individually, at each of these positions (Sambrook *et al.*, 1989; Allen *et al.*, 1998). However, since the effect of an alteration can not be

predicted with certainty, maximally 10 amino acids will be targeted for conversion to cysteine, and each variant will be assayed for activity (described below). DNA will be purified from isolated colonies, sequenced using dye-terminator fluorescent chemistry, the reaction products will be detected on an ABI PRISM 377 Automated Sequencer, and analyzed using Sequencher™ (GeneCodes, Inc.).

EXPRESSION AND PURIFICATION OF ENZYME VARIANTS

Taq polymerase mutants optimized for single-molecule sequencing will be expressed in *E. coli* from constructs created in the Hardin laboratory. Protein for experimental purposes will be prepared in the Willson laboratory, as follows.

While we have experience in growing *E. coli* to optical densities exceeding 100 by computer-controlled feedback-based supply of non-fermentative substrates, the resulting three kg of *E. coli* cell paste will be excessive for most polymerase variants, which will be of only transient interest as we engineer the polymerase to higher and higher levels of performance. We will likely apply this strategy, however, to polymerases to be used extensively in development of imaging and sequencing protocols. More commonly, we will prepare cell mass in 10 L well-oxygenated batch cultures using a rich medium designed for such purposes at Amgen. It is possible that some mutants will be prepared in 2 L baffled shake glasses. Cell paste will be harvested using our existing 6 L preparative centrifuge, lysed by French press, and cleared of cell debris by centrifugation. Because the polymerase protein will be used in DNA sequencing experiments, nucleic acid removal is desirable. Removal will be achieved using either nucleases (and subsequent heat denaturation of the nuclease) or, more likely, we will employ a variation of our compaction agent-based nucleic acid precipitation protocol [redacted]

Purification of *Taq* polymerase away from contaminating proteins will take advantage of the enormous thermal stability of this molecule relative to typical *E. coli* proteins. Heat treatment at 75°C for 60 min reduces *E. coli* protein contamination by approximately 100-fold, which when combined with the high initial expression level produces nearly pure *Taq* polymerase in a convenient initial step. (This step may not always be available for highly-engineered polymerases of reduced stability, for which we would employ more conventional techniques).

For routine sequencing and PCR purposes, limited further purification is required. A single anion-exchange step, typically on Q Sepharose at pH 8.0, has been found to suffice. It is anticipated that we will perform a second step on many samples to insure that contamination does not cloud the results of our subsequent work. Purified proteins will be characterized by SDS-PAGE and CD-monitored melting experiments where appropriate, before being passed on to the Hardin and Tu laboratories for enzymological and sequencing characterization.

POLYMERASE ACTIVITY ASSAYS USING A FLUORESCENTLY-TAGGED ENZYME AND/OR dNTP(S)

We will monitor the activity of polymerase variants throughout enzyme development. Enzyme activity will be assayed after a candidate amino acid is mutated to cysteine and following fluorescent tagging of that cysteine. A similar assay will be used to monitor the ability of a polymerase or a polymerase variant to incorporate fluorescently-tagged dNTPs. Since the enzyme's amino acid sequence will be altered, we will determine whether enzyme characteristics are altered (thermostability, fidelity, polymerization rate, affinity for modified versus natural bases). Similar procedures will be used to identify the optimal reaction buffer.

Polymerase activity assays will use conditions similar to those developed to examine single base incorporation by a fluorescently-tagged Klenow fragment DNA polymerase (Furey *et al.*, 1998). This polymerase demonstrates that the addition of a fluorescent tag does not necessarily adversely affect enzyme activity. To examine polymerase activity, the purified *Taq* will be incubated in polymerase reaction buffer with a 5'-^{32}P end-labeled primer/single-stranded template duplex, and appropriate dNTP(s). The polymerase's ability to incorporate a fluorescently-tagged dNTP will be monitored by assaying the relative amount of fluorescence associated with the extended primer on either the ABI377 DNA Sequencer or the [redacted] It will be important to demonstrate that the polymerase can continue extension following incorporation of a uniquely-tagged fluorescent dNTP. This will be assayed using an end-labeled primer, the fluorescently-tagged dNTP and the appropriate base beyond the fluorescent tag. The products will be size separated and analyzed for extension. Reactions will be either held at constant temperature or thermocycled, as necessary.

Initially, four oligonucleotides will serve as DNA templates, and each will differ only in the identity of the first base incorporated. This will allow us to examine relative incorporation efficiency of each base. Subsequently, our studies will use relatively simple-sequence, single-stranded DNA templates. A wide array of sequence-characterized templates is available in the Hardin laboratory, including a resource of over 300 purified templates [redacted] Jones and Hardin, 1998a; Jones and Hardin, 1998b; Hardin *et al.*, 1996). As an example, one series of [redacted] defined-sequence templates will be constructed as necessary, and will facilitate development of the base-calling algorithm (discussed below).

Fluorescent Tag Choice and Addition

Approach 1

The following principles will be guiding our search for appropriate fluorescence dyes for this work. In one approach, a fluorescence donor will be attached to the polymerase and four unique fluorescence acceptors will each be attached to a different dNTP. The absorption spectra of the donor and acceptor fluorophores should be sufficiently distinct to allow exclusive (preferably) or preferential excitation of the fluorescence donor attached on the polymerase at a chosen wavelength. The emission of the fluorescence donor should have significant overlap with the absorption spectra of the fluorescence acceptors. The four fluorescence acceptors, in the dNTP-attached forms, should each have a unique fluorescence emission distinguishable from that of the other three.

Several sites on dNTPs will be explored for the attachment of the fluorescence acceptors. The initial efforts will be directed to the tagging of the terminal phosphate of dNTP. This approach has a unique advantage. When the incoming, tagged dNTP is bound to the active site of the polymerase, significant FRET from the donor on the polymerase to the acceptor on the dNTP is expected to occur. The unique fluorescence of the acceptor then enables the determination of the identity of the dNTP. Once the tagged dNTP is processed for covalent attachment to the nascent DNA chain, the fluorescence acceptor remains attached to the pyrophosphate and will be released to the medium. In fact, the growing nascent DNA chain will contain only the normal dNMP building units and no fluorescence acceptor molecules at all. In essence, FRET will only occur between the donor on the polymerase and incoming acceptor-labeled dNTP, one at a time. This approach is better than the

alternative attachment of the acceptor to any site within the dNMP moiety of the initial dNTPs. In this latter case, the nascent DNA chain will contain multiple molecules of the fluorescence acceptors. Interference with the polymerase reaction and FRET measurements could occur.

Approach 2

The second approach is to only label the polymerase with a fluorophore. The dNTPs will each be labeled with a quencher of the fluorophore. Ideally, each quencher, when brought to close vicinity to the fluorophore, should have a unique quenching efficiency distinguishable from those by the other quenchers. Therefore, the degrees of quenching will allow the determination of each incoming, labeled dNTP. Four quenchers with distinguishable degrees of efficiencies may not be easy to obtain. Even with only two suitable quenchers, one can label two of the four types of dNTP with the quenchers for one run of the reaction and repeat the same DNA polymerization reaction several times, each time with a different pair of the labeled dNTPs. Results, when taken together, will enable us to definitively determine the complete sequence of the DNA molecule. One obvious advantage of this approach is that fluorescence emission will be coming from a single source. Background noise will be negligible.

General Considerations and Labeling

The fluorescence probes (or quenchers) chosen for attachment to the polymerase or dNTPs should not have any marked adverse effects on the DNA polymerization reaction. The rational and strategies for the selection of sites on the polymerase and dNTPs for the probe attachment are described elsewhere in this proposal. Procedures will be developed to chemically tag the polymerase and dNTPs with the chosen fluorescence probes (or quenchers). In general, the polymerase with specific residue(s) targeted or created mutationally for labeling will be treated with a slight molar excess of a desired probe in the hope that near stoichiometric labeling can be achieved. Alternatively, the polymerase can be treated with an excess amount of the probe and the labeling will be followed as a function of time. The tagging reaction will be stopped when near stoichiometric labeling is obtained. The possibility that excessive tagging of residues other than the targeted one occurs leading to adverse effects on enzyme activity or subsequent FRET measurement should be considered. If the targeted residue is close to the active site, a saturating level of substrate or a competitive inhibitor can first be added to protect the targeted residue at the enzyme active site and a reversible labeling reagent can be subsequently added to tag these non-active site residues. The modified enzyme will be freed from the protective substrate (or competitive inhibitor) and remaining free reversible reagent, and treated with the desired fluorescence probe for the labeling of the targeted residue. Finally, the reversible tags will be chemically freed from the enzyme and removed to obtain the polymerase containing the desired fluorescence donor attached to primarily the targeted residue. Alternatively, the targeted residue may not be near the active site. Excessive labeling of other residues could only occur if they are significantly more reactive than the targeted residue for tagging. The polymerase can also be treated with a reversible reagent for preferential labeling of those residues which are not selected for fluorescence probe attachment, but are chemically more susceptible for tagging. After removal of the remaining free reversible reagent, the modified enzyme can then be treated with the desired fluorescence probe for the labeling of the less reactive targeted residue. Finally, the reversible tags can be chemically freed from the enzyme and removed to obtain the polymerase with the fluorescence probe attached to primarily the targeted residue.

Dr. Tu is experienced in various aspects of enzymology (including chemical modification and site-specific labeling of enzymes) and fluorescence spectroscopy. Selected relevant publications from his work are included in his curriculum vitae.

CHEMICAL MODIFICATION OF NUCLEOTIDES FOR DNA POLYMERASE REACTIONS

Specific Aims

- Develop synthesis of modified fluorophore and fluorescence energy transfer compounds of distinctly optical properties for differential signal detection.
- Develop synthesis of nucleoside/nucleotide synthons for incorporation of modifications on base, sugar or phosphate backbone positions.
- Develop synthesis of complementary sets of four deoxynucleotide triphosphates (dNTPs) containing substituents on nucleobases, sugar or phosphate backbone.

Significance

A high level incorporation of dNTPs is crucial for the success of the proposed project. It is, thus, requiring intense chemistry effort in synthesis of modified dNTPs to engineer features that would permit high fidelity enzymatic synthesis and sensitive detection. Presently, a number of dye-labeled dNTPs are available from commercial resources. However, the protein-DNA complex system used in our method imposes demands that are more stringent, such as null background signals with minimal interference in multi-fluorophore systems. These requirements cannot be completely satisfied by commercial products.

Research Plan

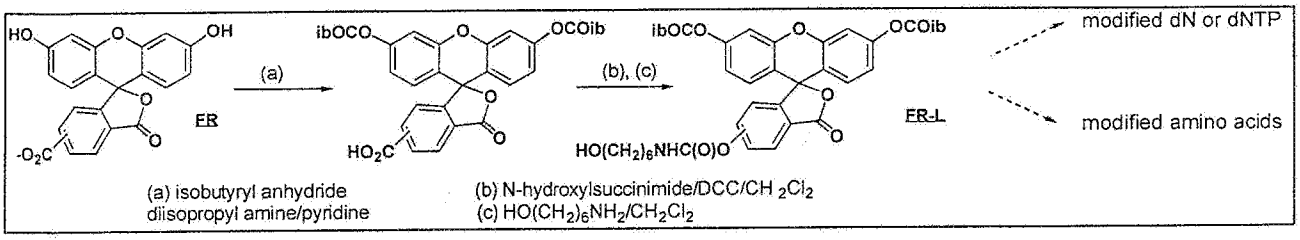
The proposed synthesis will be based on the nucleoside/nucleotide chemistry developed in this laboratory as part of the antisense oligonucleotide (AON) project. In the AON project, our interests are to understand the correlation of chemical structure modifications with AON binding affinity and specificity in target sequences. Using chemistry and high resolution NMR in combination, our laboratory has characterized a series of AONs (Gao *et al.*, 1992; Rice and Gao, 1997; Cross *et al.*, 1997; Gao *et al.*, 1997[redacted]. The chemistry of modified nucleotides used in the AON project is directly applicable for the proposed synthesis. This background would permit us proceed rapidly to achieve synthesis of modified dNTPs.

In the proposed project, we will work closely with Dr. Tu in selection of molecules for signal detection. We initially choose to use the popular fluorescing molecules, such as rhodamine and fluorescein derivatives, and utilize the fluorescence resonance energy transfer (FRET) phenomenon (Foster, 1965; Ju *et al.*, 1995; Lee et al, 1997; Furey *et al.*, 1998). Alternatively, chromophore interactions as in a fluorophore-quencer pair (Tyagi and Kramer, 1996; Tyagi *et al.*, 1998[redacted]) or a fluorophore-excimer pair (Yamana *et al.*, 1997; Tong *et al.*, 1995; Paris *et al.*, 1998; Lewis *et al.*, 1997) may be considered. Together, these molecules are called tags. In these designs, we would need to place a tag on the polymerase and its energy partner tag on the dNTP. The choice of fluorophore is a function of not only its enzyme compatibility, but also its spectral and photophysical properties. For instance, it is critical that the acceptor fluorophore does not have absorption (*i.e.*, at least less than 1/1000) at the excitation wavelength of the donor fluorophore, and that the donor fluorophore does not have emission at the detection wavelength of the acceptor fluorophore. These spectral characteristics may be attenuated by chemical modifications of the fluorophore ring systems. Absorbance and emission spectra of the modified fluorescing molecules will be examined to satisfy the requirements discussed above.

In the following, we provide reaction routes that serve as examples for the proposed synthesis. These synthesis reactions have been used in our on-going projects in the AON area and DNA microarrays and demonstrate our current effort and capability for developing the chemistry to meet the demand of the proposed project.

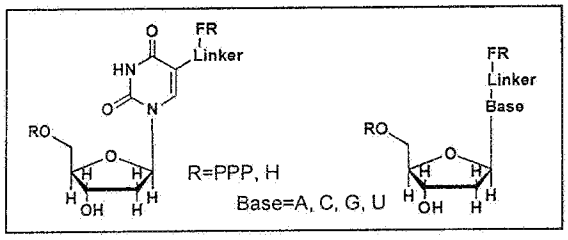
Synthesis of fluorescein derivatives. Fluorescein (FR) molecules will be modified to contain a linker unit. These molecules can be covalent attached to nucleotides (Ward *et al.,* 1987; Engelhardt *et al.,* 1993; Hobbs, 1991) or amino acids. A representative synthesis is shown below (Scheme 1). The product FR-L can be used to attach to nucleotides and amino acids. Other fluorophore molecules may be modified using similar type of chemistry.

Scheme 1



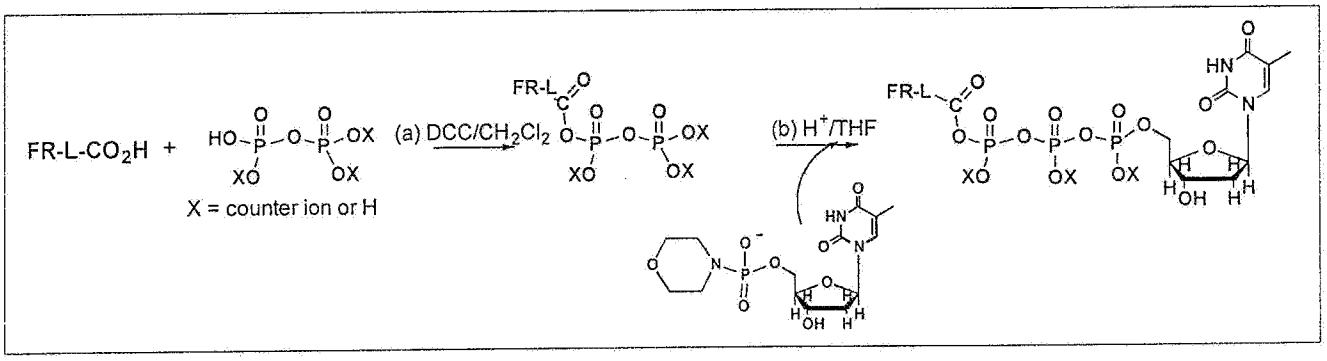
Synthesis of base and sugar modified nucleotide dU. Nucleobases and sugar moieties can be modified with a fluorophore, yet still maintain their enzymatic reaction activity. The modifications are also selected as sites that do not interference with Watson-Crick base pairing. The basic structural scheme for base modification is shown in Scheme 2. Our laboratory routinely prepares nucleotide derivatives in milligram quantities and has procedures for preparation of tagged nucleotides, which are not commercially available.

Scheme 2



Synthesis of modified dNTPs. We hypothesize that polymerase may be able to utilize phosphate-modified dNTPs. If a tagged _-phosphate ester can be used as a substrate, then the tag will be removed by the enzyme after nucleotide incorporation. Since the replicated DNA will not contain any unnatural bases, polymerase activity is less likely to be affected and extended strands should result. We will synthesize phosphate-modified nucleotides using adapted literature procedures (Bonnaffe *et al.,* 1995). An example for the reaction of _-phosphate modification is shown in Scheme 3.

Scheme 3



Our synthesis will initially focus on pyrimidine nucleotides and identify suitable tags. Effort will also be made to change the substituents on fluorophore and relevant molecules. These chemical conversions may be necessary for achieving sufficient levels of incorporation by the polymerase. Additionally, it is anticipated that multicolor (or intensity) detection will improve confidence values associated with the base calling algorithm. These compounds will be tested initially in our laboratory

and then by the project collaborators. Chemistry will be constantly revised according to input from these laboratories.

DEVELOPMENT OF A SINGLE-MOLECULE DETECTION SYSTEM

Single-molecule florescence imaging will employ our existing research-grade Nikon Diaphot TMD inverted epifluorescence microscope, upgraded with laser illumination and a more-sensitive camera. While we have no direct experience in single-molecule fluorescence detection, the literature abounds with references of single-molecule detection (Goodwin *et al.*, 1997; Ambrose *et al.*, 1994; [redacted]; Castro and Williams, 1997; Keller *et al.*, 1996; [redacted]; Davis *et al.*, 1992; Orrit and Bernard, 1990; Orrit *et al.*, 1994; Davis *et al.*, 1991). Additionally, we do have experience with all of the required techniques, including fluorescence spectroscopy, evanescent wave-illumination, CCD-based digital imaging and image processing, and imaging-based monitoring of catalyst and enzyme activity in groundbreaking new devices. We also have available on the campus

[redacted]

microscope will be retrofitted for evanescent-wave excitation using an argon ion laser at 488 nm. We have previously used this illumination geometry in assays for nucleic acid hybridization. The other major modification to our existing setup is replacement of the current CCD camera with a 12-bit 512 x 512 pixel Princeton Instruments I-PentaMAX generation IV intensified CCD camera, which has been used successfully in a variety of similar single-molecule applications. This camera achieves a quantum efficiency of over 45% in the entire range of emission wavelengths of the dyes to be used, and considerably beyond this range. The vertical alignment of our existing microscope will tend to minimize vibration problems, and the instrument is currently mounted on an anti-vibration table.

We will approach the development of a fully-functional, four-color, real-time, single-molecule imaging system in stages. Our original setup for demonstration purposes will follow only one emission wavelength at a time, to minimize instrument complexity and cost, and to facilitate proof of concept as rapidly as possible. For reasons discussed elsewhere in this proposal, it may be necessary to convolve partial information obtained from multiple polymerase molecules in order to determine the overall sequence of the template molecule. Given this constraint, single-color detection is not a major handicap for the near- and medium-term development of the technique. An important driving force for convolving together results obtained with multiple single-molecules is the impossibility of obtaining data from a single molecule over an indefinite period of time. At a typical dye photobleaching efficiency of $2*10^{-5}$ a typical dye molecule would be expected to undergo 50,000 excitation/emission cycles before permanent photobleaching. Data collection from a given molecule may also be interrupted by intersystem crossing to an optically inactive (on the time scales of interest) triplet state. Even with precautions against photobleaching, therefore, data obtained from any given molecule will necessarily be fragmentary for template sequences of substantial length, and it is necessary to plan for

[redacted]

Polymerase Activity Assays Using a Single-Molecule Detection System

These assays will be performed essentially as described in the "Polymerase Activity Assays Using a Fluorescently-tagged Enzyme and/or dNTP(s)" section. The primary difference involves the immobilization of either the polymerase or the DNA to a solid support to enable viewing of an individual replication event. A variety of immobilization options will be investigated, including immobilization on a silica surface (Basche *et al.*, 1992; Ambrose *et al.*, 1994).

Analysis of fluorescent signals from single molecule sequencing system

The raw data generated by the detector will represent between one to four time-dependent data streams of fluorescence wavelengths and intensities, one data stream for each fluorescently labeled base (i.e. wavelength) being monitored. We will initially attempt to use the PHRED computer program (Ewing *et al.*, 1998) to assign base identities and reliabilities. If needed, we will write computer programs to interpret the data streams. As mentioned above, we will need to piece together partial and overlapping sequences. Multiple experiments will be run so that confidence limits can be assigned to each base identity according to the variation in the reliability indices and the difficulties associated with assembling stretches of sequence from fragments. The reliability indices represent the goodness of the fit between the observed wavelengths and intensities of fluorescence compared with the ideal values. The result of the signal analyses is a linear DNA sequence with associated probabilities of certainty.

Project Milestones Toward Real-Time Sequence Determination (Including Long-Range Plans)

- Overall coordination of project efforts (Hardin) active communication and optimization of interdependencies of the project (Hardin, Briggs, Gao, Tu, Willson).
- We will identify the most appropriate polymerase for our studies (Hardin).
- Structural modeling and interpretation of results identify the optimal modification site for tag attachment on the (Briggs, Hardin).
- We will engineer, express, and purify the polymerase (Hardin).
- Fermentation, purification, and quality control of polymerase candidate protein (Willson).
- We will identify and/or design optimal fluorescence dyes for the attachment to the engineered polymerase (Gao, Tu).
- Biochemical, enzymological, and sequencing performance characterization of novel polymerases (Willson, Tu, Hardin).
- Molecular modeling will provide a rational method for identifying the best sites for fluorescently labeling the dNTP, and for estimating fluorescence resonance energy transfer (FRET) efficiency (Briggs, Gao, Tu).
- We will identify and/or design optimal labeling system (Gao, Tu).
- Characterization of FRET behavior of labeled polymerase/labeled substrate systems (Tu, Willson)
- Design of detection systems. One will be used to assay progress in enzyme design and modification using non-challenging detection methods. The second will be a prototype that will unite technologies and enable single molecule detection. Construction and optimization of single-molecule fluorescence apparatus and techniques (Willson, Tu)
- We will develop a computer algorithm to interpret the fluorescent signals and present the user with DNA sequence information (Briggs, Hardin).
- Initial demonstration of sequencing past one fluorescently-labeled base inserted into a sequence of unlabelled DNA. Detection of products by electrophoresis (Hardin, Tu).

- Demonstration of activity of labeled polymerase with unlabeled substrate (Hardin, Tu).
- Detection of the product using our existing ABI sequencer (Hardin).
- Demonstration of the ability of a labeled polymerase to read past one labeled base of each type (beginning with one particular type, and then by iteration and further improvement extended to all four types). Product detection by electrophoresis (Hardin, Tu).
- Detection of resonance energy transfer between labeled polymerase and labeled substrate molecules on an ensemble basis using our existing SPEX 212 fluorometer with thermoelectrically-cooled PMT (Willson, Tu)
- Completion of single-molecule fluorescence upgrade to our existing epifluorescence microscope by addition of ICCD camera and laser excitation (Willson).
- Observation of single-molecule florescence signals from labeled polymerase acting on a substrate in which one of the four bases is uniformly labeled (Willson, Hardin, Tu).
- Extension of the previous result to detection of florescence signals from polymerases acting on each of the other three types of single-base-type-labeled substrates (with polymerases individually optimized for read through and signal intensity from the individual types of base, not necessarily all the same type of polymerase). (Willson, Hardin, Tu)
- Observation of single-molecule florescence signals from polymerases acting on one-base-type-labeled substrates to give signals that correlate with the known sequence of the substrate, and extension of this result to all four types of one-base-type-labeled substrate (Willson, Hardin, Tu).
- Development of informatics tools for assembling partial sequence information obtained from single-molecule experiments into complete sequence with at least 99% accuracy. (Requires compensation for termination due to photobleaching). (Hardin, Briggs)
- The dsDNA with be treated with DNaseI to randomly introduce nick(s). The 3' end of a nick will serve as the primer for extension. A laser will stimulate fluorescence once the polymerase tracks (translocates) to the site of the nick. Once at the nick the polymerase will perform a nick translation reaction, initiating the sequence reaction. This method requires no sequence knowledge of the genome. (Willson, Tu, Hardin)
- Development of labeled polymerase able to read through and produce single-molecule florescence data from two, then three, then four-base-type-labeled substrates. (Willson, Hardin, Tu)
- Automation of single-molecule florescence sequencing apparatus by addition of sample-handling and microfluidics front end. (Willson, Hardin, Tu, Gao)
- Miniaturization of single-molecule sequencing apparatus for clinical and laboratory applications (Willson, Hardin, Tu, Gao, Briggs).

INFORMATION AND TECHNOLOGY TRANSITION

All of the investigators are active participants at a variety of national and international scientific meetings. The PI (Hardin) has developed a primer-walking sequencing strategy using

The project leaders, postdoctoral fellows, graduate students, and undergraduate students will participate in monthly meetings to discuss progress made throughout the award period. Informal, discussion-promoting presentations will be made by each group at these meetings. This will foster cross-disciplinary training for *all* personnel involved in the project.

CONTINUATION PLANS

Biosciences, biotechnology, and bioengineering are areas of major commitment by the

Please see attached letter from Dr. Arthur Vailas, Vice President for Research, University of Houston.

REFERENCES

Allen, B. L., F. Uhlmann, L. K. Gaur, B. A. Mulder, K. L. Posey, L. B. Jones, S. H. Hardin. 1998. DNA recognition properties of the N-terminal DNA binding domain within the large subunit of Replication Factor C. *Nucleic Acids Research* **26**, 3877-3882.

Ambrose, WP, PM Goodwin, JC Martin, and RA Keller. 1994. Alterations of single molecule fluorescence lifetimes in near-field optical microscopy. *Science* 265, 364-367.

Ball, S., Reeve, M.A., Robinson, P.S., Hill, F., Brown, D.M., and Loakes, D. 1998. The use of tailed octamer primers for cycle sequencing. *Nucleic Acids Research* **26**;5225-5227.

Basche T, Moerner WE, Orrit M, Talon H. 1992. Photon antibunching in the fluorescence of a single dye molecule trapped in a solid. *Phys. Rev. Lett.* 7;69(10):1516-1519.

Bedford E, Tabor S, Richardson CC. 1997. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci. USA* 94(2),479-484.

Bonnaffe, D., Dupraz, B., Ughetto-Monfrin, J., Namane, A., and Dinh, T. H. (1995) Synthesis of acyl pyrophosphates. Application to the synthesis of nucleotide liphophilic prodrugs. *Tetrahedron Lett.* 36, 531-534.

Burbelo, P. D. and Iadarola, M. J. 1994. Rapid Plasmid DNA Sequencing with Multiple Octamer Primers. *BioTechniques* **16**;645-650.

Castro, A., and JGK Williams. 1997. Single-molecule detection of specific nucleic acid sequences in unamplified genomic DNA. *Anal. Chem* 69, 3915-3920.

Clegg, R. M., Murchie, A. I. H., Zechel, A., Lilley, D. 1993. Observing the Helical Geometry of Double-Stranded DNA in Solution by Fluorescence Resonance Energy Transfer. *Proc. Natl. Acad. Sci. USA* **90**; 2994-2998.

Cline, J, JC Braman, and HH Hogrefe. 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA plymerases. *Nucleic Acids Research* 24, 3546-3551.

Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., the members of the DOE and NIH planning groups. 1998. New Goals for the U.S. Human Genome Project: 1998-2003. *Science*. **282**: 682-689.

Cross, C. W., Rice, J. S., and Gao, X. (1997) Solution structure of a RNA·DNA hybrid duplex containing a 3'-thioformacetal linker and an RNA A-tract. *Biochemistry* 36, 4096-4107.

Davis LM, Fairfield FR, Harger CA, Jett JH, Keller RA, Hahn JH, Krakowski LA, Marrone BL, Martin JC, Nutter HL, et al. 1991. Rapid DNA sequencing based upon single molecule detection. Genet Anal Tech Appl. 1991 8(1):1-7.

Davis, L., FR Fairfield, ML Hammond, CA Harger, JH Jett, RA Keller, JH Hahn, LS Krakowski, B Marrone, JC Martin, HL Nutter, RR Ratliff, EB Shera, DS Simpson, SA Soper, and CW Wilkerson. 1992. Rapid DNA Sequencing based on single-molecule detection. *Los Alamos Science* 20, 281-285.

Eckert, KA and TA Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research* 11, 3739-3744.

Engelhardt, D., Rabbani, E., Kline, S., Stavrianopoulos, J. G., and Dollie, K. (1993) Base moiety-labeled detectable nucleotide. Patent, US5241060.

Eom, SH, J Wang and TA Steitz. 1996. Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* 382, 278-281.

Ewing, B., Green, P. 1998. Base-Calling of Automated Sequencer Traces using PHRED. II. Error Probabilities. *Genome Res.* **8**; 186-194. Ewing, B., Hillier, J., Wendl, M. C., Green, P. 1998. Base-Calling of Automated Sequencer Traces using PHRED. I. Accuracy Assessment. *Genome Res.* **8**; 175-185.

Foster, T. (1965) Modern quantum chemistry. Istanbul Lectures, Part III, Academic: New York. Pp.93-137.

Furey, W. S., Joyce, C. M., Osborne, M. A., Klenerman, D., Peliska, J. A., Balasubramanian, S. 1998 Use of Flourescence Resonance Energy Transfer to Investigate the Conformation of DNA Substrates Bound to the Klenow Fragment. *Biochem.* **37**; 2979-2990.

Gao, X., Brown, F. K., Jeffs, P., Bischofberger, N., Lin, K-Y., Pipe, A. J., and Noble, S. A. (1992) Probing structural factors stabilizing antisense oligonucleotide duplexes. NMR studies of DNA•DNA duplexes containing a formacetal linkage. *Biochemistry* 31, 6228-6236.

Gao, X., Rice, J. S., Cross, C. W., Yang, X. (1997) Structural correlations of backbone modifications in antisense oligonucleotide duplexes. *Nucleosides and Nucleotides* 16, 1599-1608. [review].

Goodwin, PM., H Cai, JH Jett, S Ishaug-Riley, NP Machara, DJ Semin, A van Orden, RAKeller.1997.Application of single molecule detection to DNAsequencing.*Nucleosides Nucleotides* 16(5 & 6), 543-550.

Hardin, S.H., Jones, L.B., Homayouni, R., and McCollum, J.C. 1996. Octamer Primed Cycle Sequencing: Design of an Optimal Primer Library. *Genome Research* **6**;545-550.

Hobbs, F. W. (1991) Fluorescent reagents for the preparation of 5'-tagged oligonucleotides. Patent, US4997928.

Holland, PM, RD Abramson, R Watson, and DH Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88, 7276-7280.

Huber HE, Tabor S, Richardson CC. 1987. *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. *J. Biol. Chem.* 262(33), 16224-32.

Jones, L.B., and Hardin, S.H. 1998a. Octamer-Primed Cycle Sequencing Using Dye-Terminator Chemistry. *Nucleic Acids Research* **26**;2824-2826.

Jones, L.B., and Hardin, S.H. 1998b. Octamer Sequencing Technology: Optimization Using Fluorescent Chemistry. *ABRF News* **9** (2);6-10.

Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N., and Mathies, R. A. (1995) Fluorescence energy transfer dye labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci.* USA 92, 4347-4351.

Keller, RA, WP Ambrose, PM Goodwin, JH Jett, JC Martin, and M Wu. 1996. Single-molecule fluorescence analysis in solution. *Applied Spectroscopy* 50, 12A-32A.

Kelman, Z., J Hurwitz, M O'Donnell. 1998. Processivity of DNA polymerases: two mechanisms, one goal. *Structure* 6, 121-125.

Kieleczawa, J., Dunn, J. J., and Studier, F. W. (1992) DNA Sequencing by Primer Walking with Strings of Contiguous Hexamers. *Science* **258**;1787-1791.

Kotler, L. E., Zevin-Sonkin, D., Sobolev, I. A., Beskin, A. D., and Ulanovsky, L. E. 1993. DNA Sequencing: Modular Primers Assembled from a Library of Hexamers or Pentamers. *Proc. Natl. Acad. Sci. USA* **90**;4241-4245.

Lee, LG, SL Spurgeon, CR Heiner, SC Benson, BB Rosenblum, SM Menchen, RJ Graham, A Constantinescu, KG Upadhya nd JM Cassel. 1997. New energy transfer dyes for DNA sequencing. *Nucleic Acids Research* 25, 2816-2822.

Lewis, F. D., Zhang, Y., and Letsinger, R. L. (1997) Bispyrenyl excimer fluorescence: A sensitive oligonucleotide probe. *J. Am. Chem. Soc.* 119, 5451-5452.

Li, Y, S. Korolev, and G. Waksman. 1998. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* 17(24):7514-7525.

Li, Y, S. Korolev, and G. Waksman. 1998. Crystal structures of the Klenow fragment of *Thermus aquaticus* DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Science* 7, 1116-1123.

Lichtarge O., Bourne, H. R., Cohen, F. E. 1996. An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families. *J. Mol. Biol.* **257**; 342-358.

Lundberg KS, Shoemaker DD, Adams MW, Short JM, Sorge JA, Mathur EJ. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene*. 108(1):1-6.

MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B **102**; 3586-3616.

Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. 1998. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comp. Chem.* **19**: 1639-1662.

Orrit M, Bernard J. 1990. Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal. *Phys. Rev. Lett.* **19**;65(21):2716-2719.

Orrit, M., J. Bernard, R. Brown, L. Fleury, J. Wrachtrup, C. von Borczyskowski. 1994. Single molecule fluorescence: from excitation spectra to time correlation. *J. Luminescence* 60&61, 991-996.

Paris, P. L., Langenhan, J. M., and Kool, E. (1998) Probing DNA sequences in solution with a monomer-excimer fluorescence color change. *Nucleic Acids Res.* 26, 3789-3793.

Raja, M. C., Zevin-Sonkin, D., Shwartzburd, J., Rozovskaya, T. A., Sobolev, I. A., Chertkov, O., Ramanathan, V., Lvovsky, L., Ulanovsky, L. E. 1997. DNA sequencing using differential extension with nucleotide subsets (DENS). *Nucleic Acids Res* **25**:800-805.

Reeve MA and CW Fuller. 1995. A novel thermostable polymerase for DNA sequencing. *Nature* 376, 796-797.

Rice, J. S., and Gao, X. (1997) Conformation of formacetal and 3'-thioformacetal nucleotide linkers and the stability of their antisense RNA·DNA hybrid duplexes. *Biochemistry* 36, 399-411.

Rosenblum, BB, LG Lee, SL Spurgeon, SH Khan, SM Menchen, CR Heiner, and SM Chem. 1997. New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Research* 25, 4500-4504.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.

Sanger, F., Nicklen, S., Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* U S A **74**:5463-5467.

Siemieniak, D. R. and Slightom, J. L. 1990. A Library of 3342 Useful Nonamer Primers for Genome Sequencing. *Gene* **96**;121-124.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., Hood, L E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**:674-679.

Studier, F. W. 1989. A Strategy for High-Volume Sequencing of Cosmid DNAs: Random and Directed Priming with a Library of Oligonucleotides. *Proc. Natl. Acad. Sci. USA* **86**;6917-6921.

Tabor S, Huber HE, Richardson CC. 1987. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol Chem.* 262(33),16212-23.

Tabor S, Richardson CC. 1995. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. USA* 92(14), 6339-6343.

Tong, G., Lawlor, J. M., Tregear, G. W., Haralambidis, J. (1995) Oligonucleotide-Polyamide Hybrid Molecules Containing Multiple Pyrene Residues Exhibit Significant Excimer Fluorescence. *J. Am. Chem. Soc.* 117, 12151-12158.

Tyagi, S., and Kramer, F. R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotech.* 14, 303-308.

Tyagi, S., Bratu, D. P., and Kramer, F. R. (1998) Multicolor molecular beacons for allele discrimination. *Nature Biotech.* 16, 49-53.

Ward, D. C., Langer, P. R., and Waldrop, III, A. A. (1987) Modified nucleotides and methods of preparing and using the same. Patent, US4711955.

Yamana, K., Takei, M., and Nakano, H. (1997) Synthesis of oligodeoxyribonucleotide derivatives containing pyrene labeled glycerol linkers: enhanced excimer fluorescence on binding to a complementary DNA sequence. *Tetrahedron Lett.* 38, 6051-6054.

FACILITIES AVAILABLE

Major Equipment Controlled by the Project Investigators:

Major Equipment Available to the Project Investigators:

Computers:

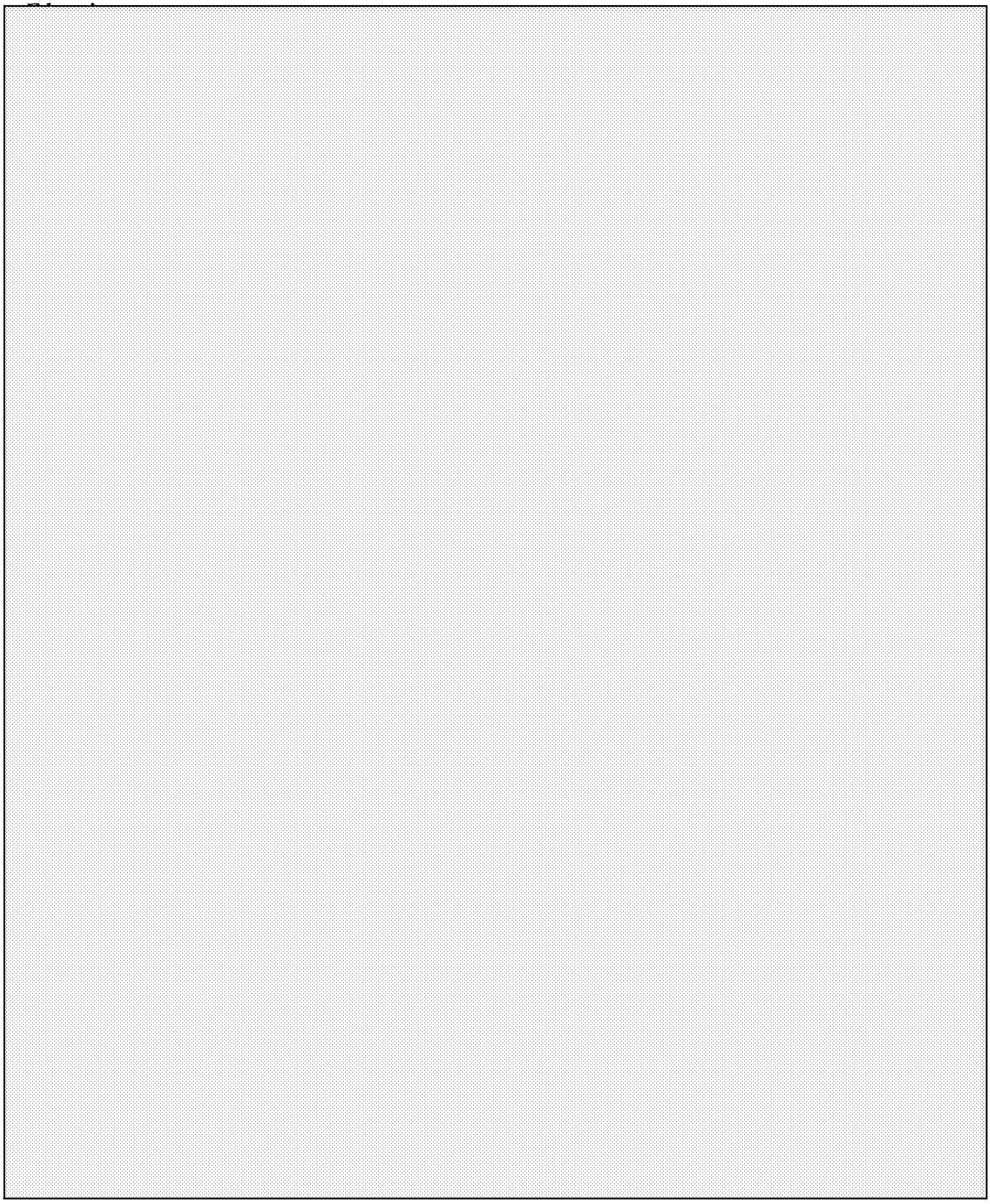
General Services:
Stockroom: Sale of enzymes, chemicals, research supplies and office/computer supplies. Machine, glass, and electronic shops in the Chemistry Department are available for our use. A darkroom with a film processor and photographic equipment is available in Science and Research 2 Building.

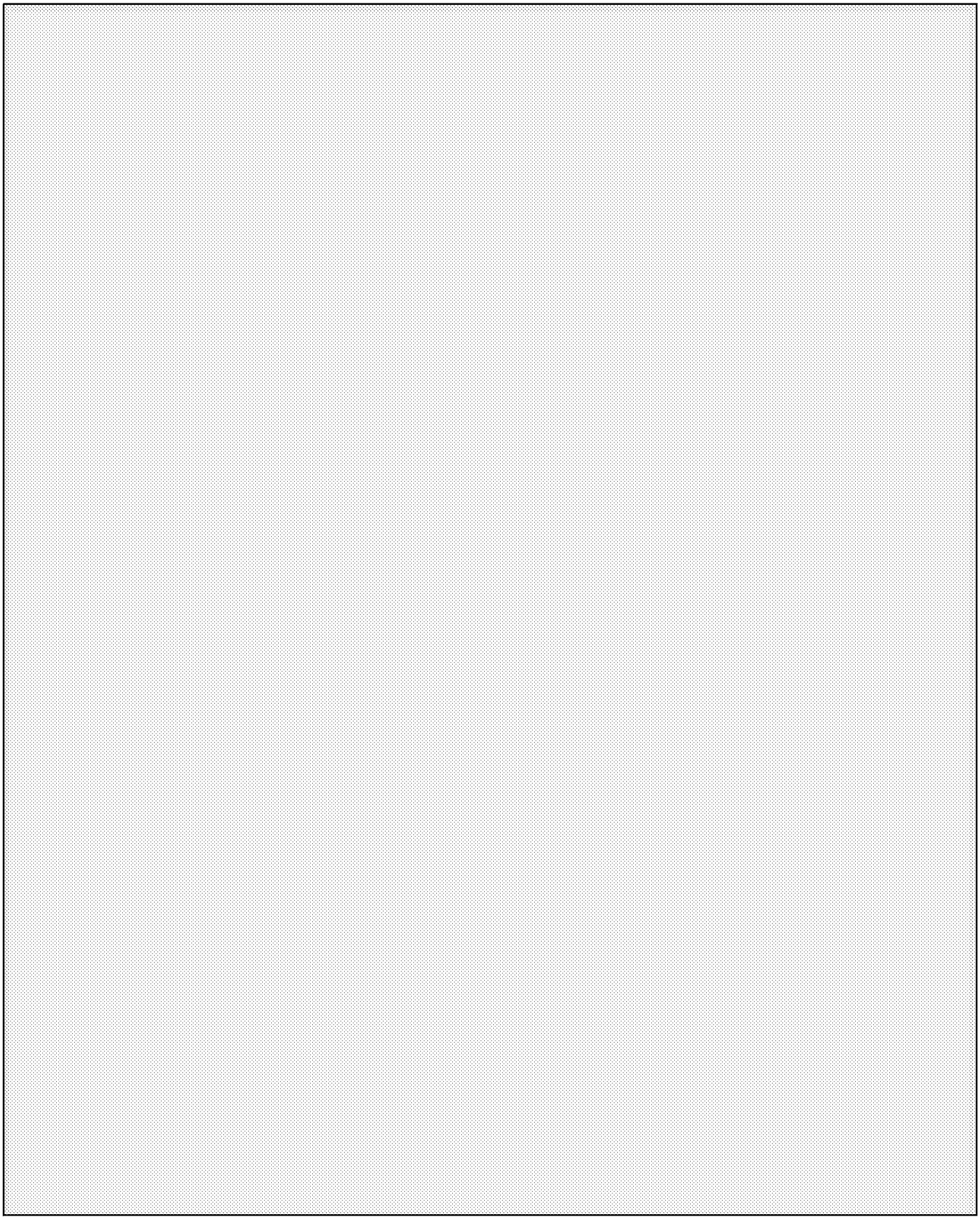
Institutes, Centers, and the Houston Scientific Community:

INSTITUTIONAL COMMITMENT TO ESTABLISH AND MAINTAIN A SUPPORTIVE INTERDISCIPLINARY ENVIRONMENT
Please see the letter of support from Dr. Arthur Vailas, Vice President of Research at University of Houston. Additionally, the following highlights a University support of our efforts.

CURRICULUM VITAE

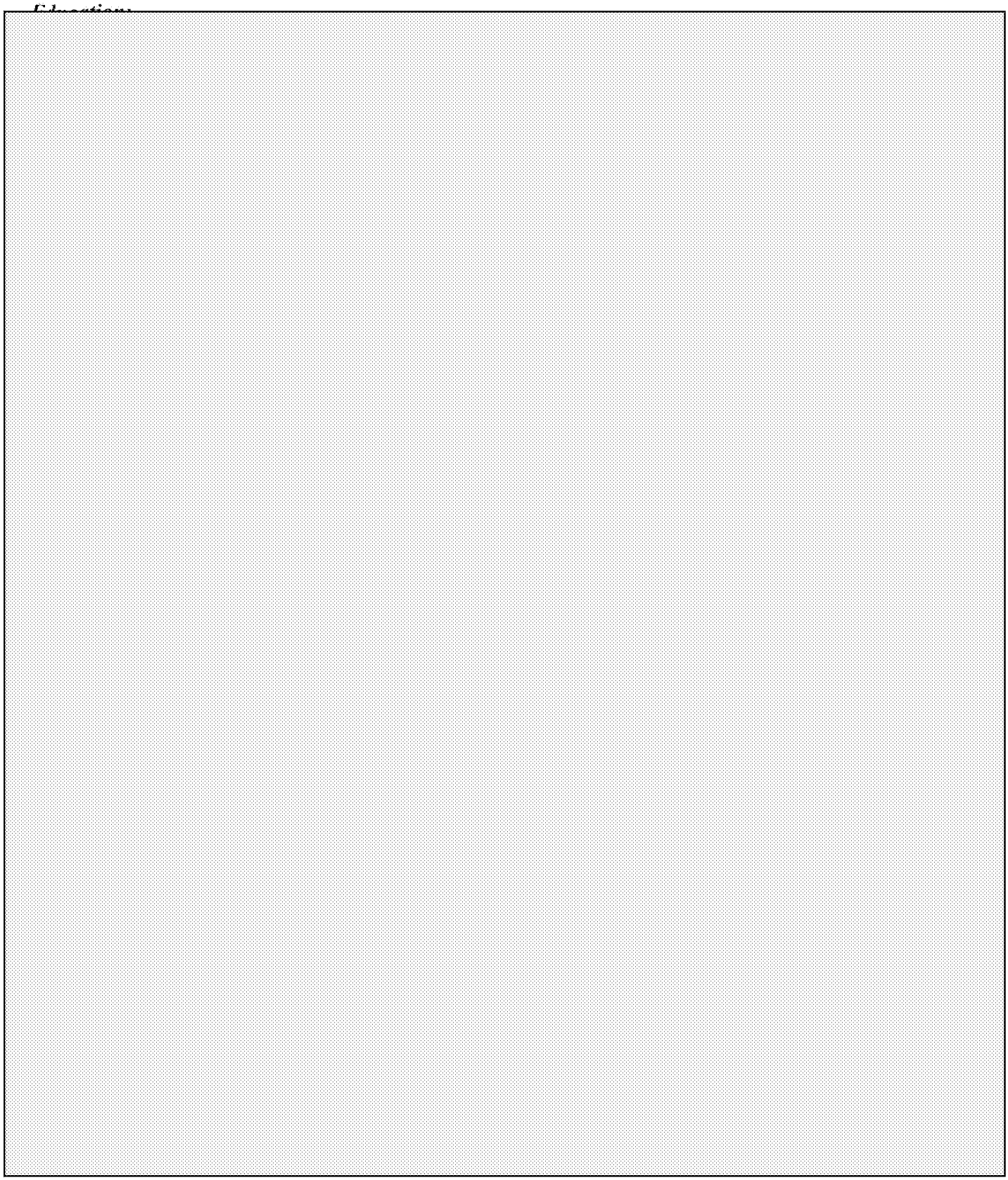
SUSAN HOUCK HARDIN

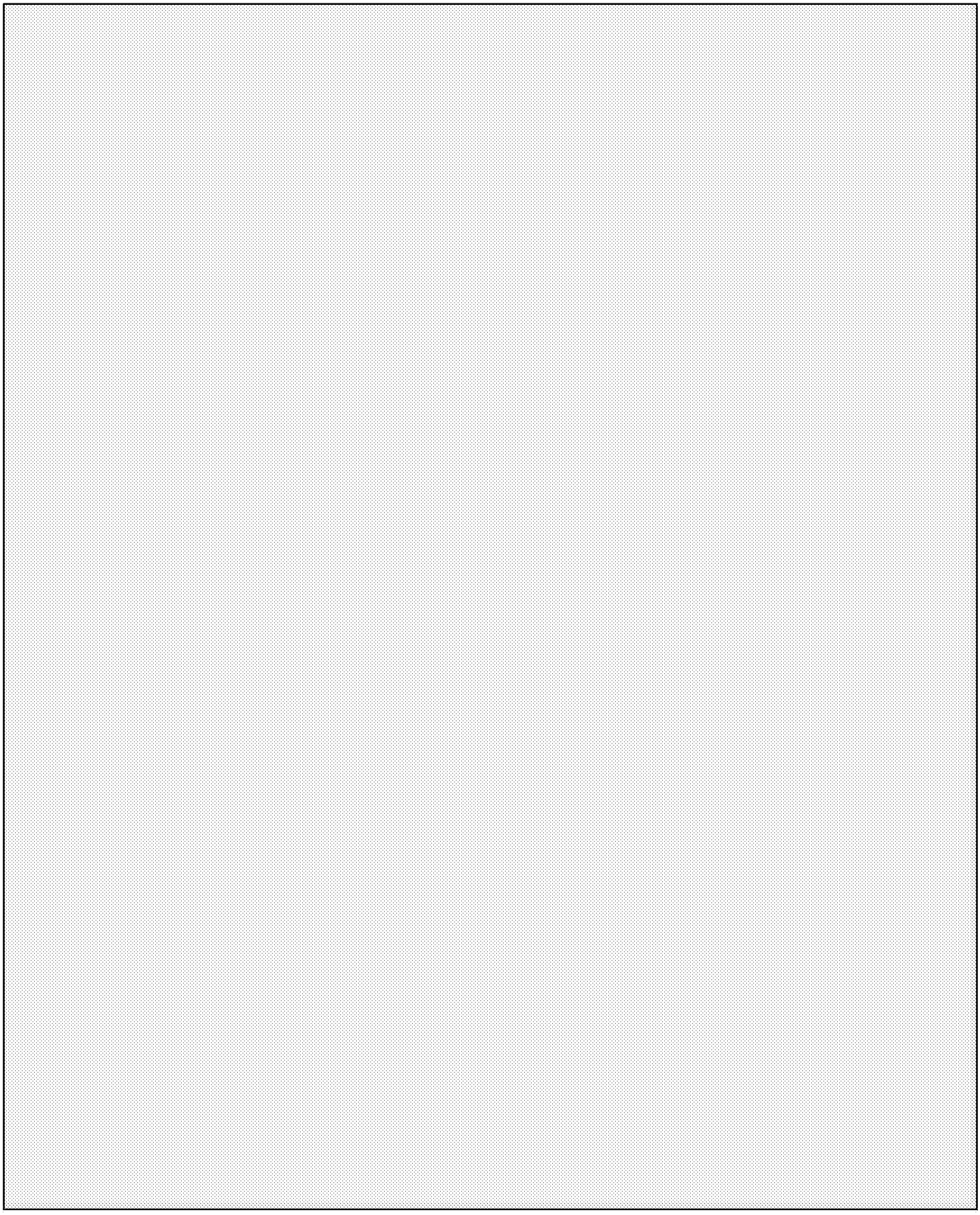




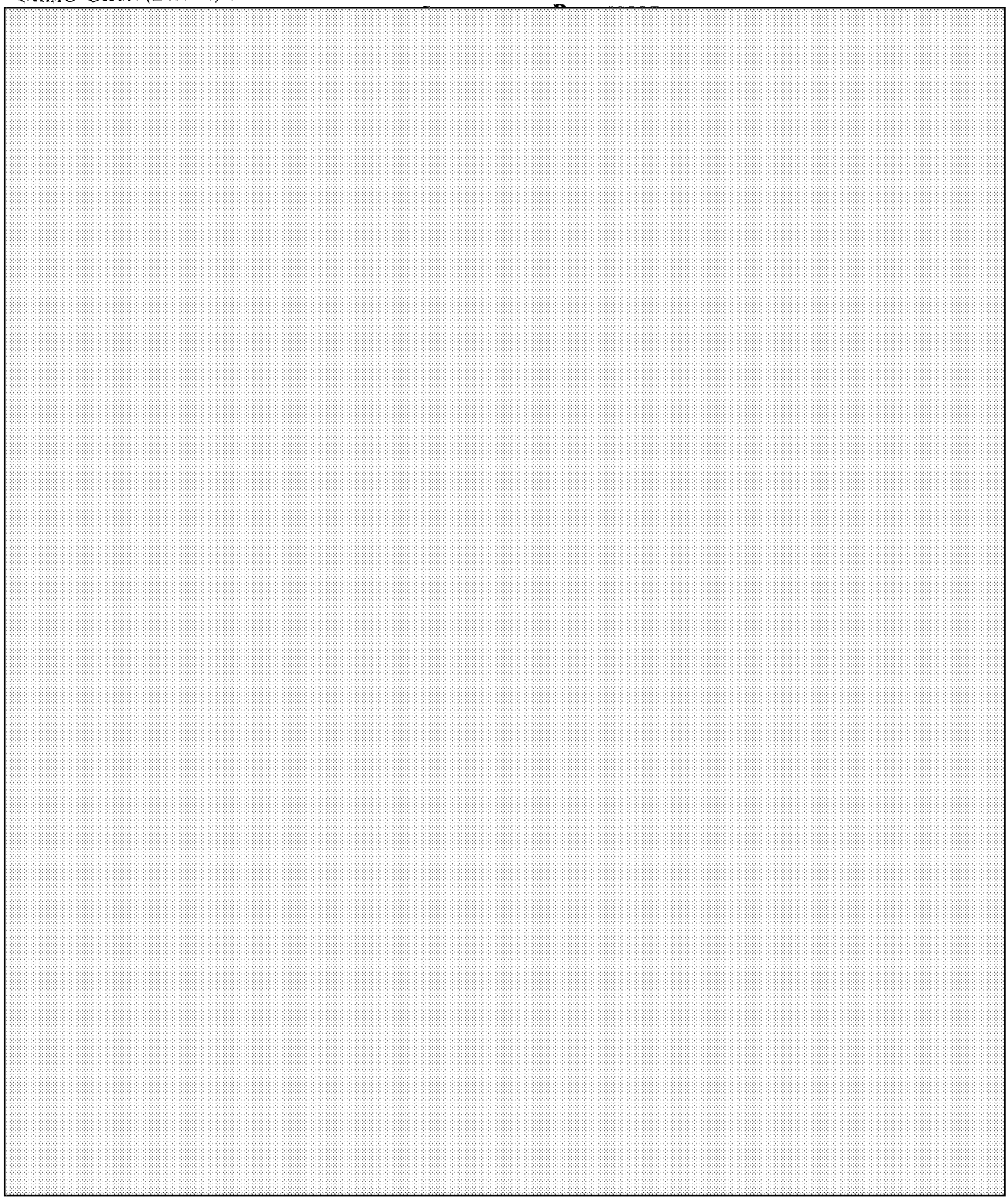
JAMES M. BRIGGS

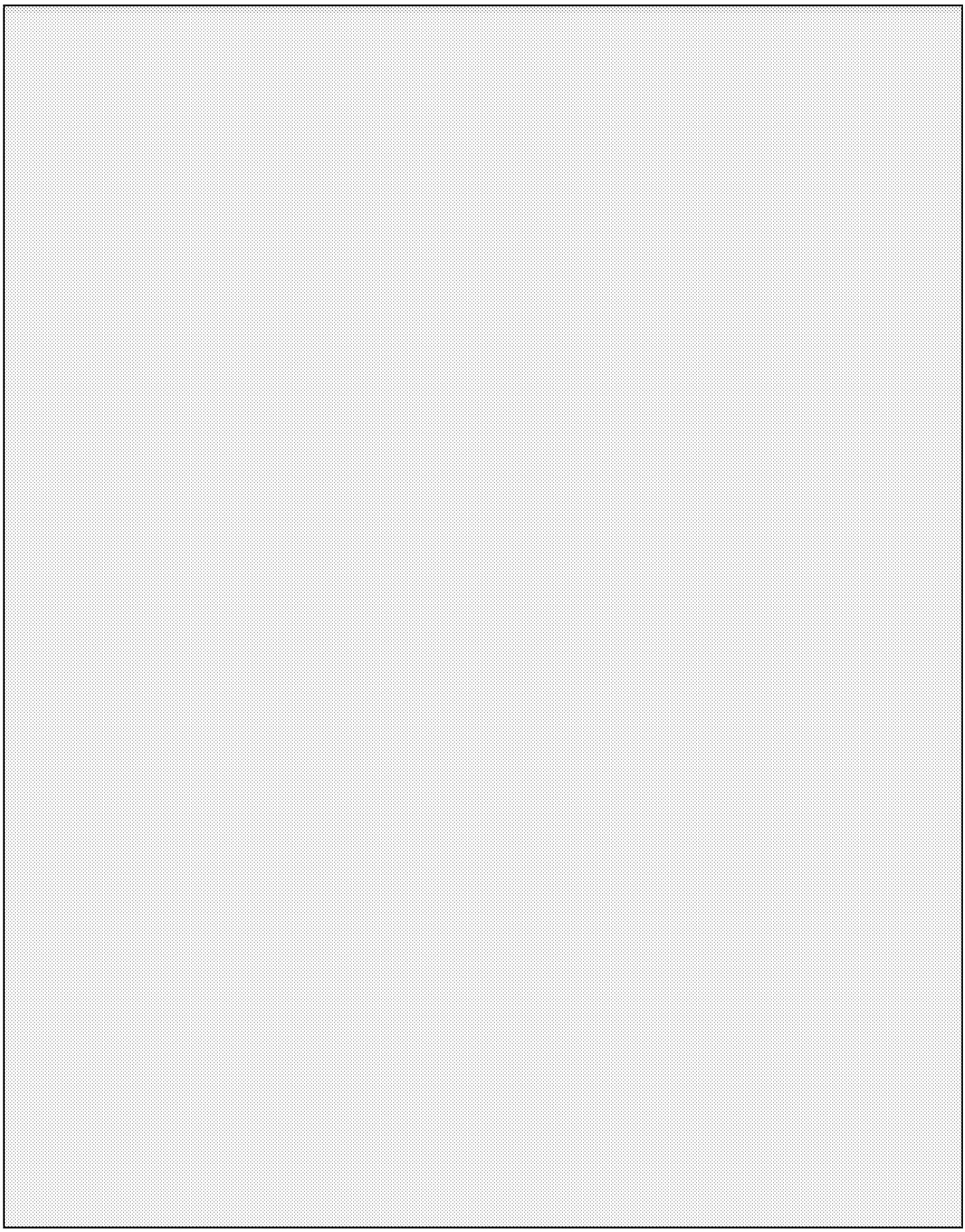
Education:

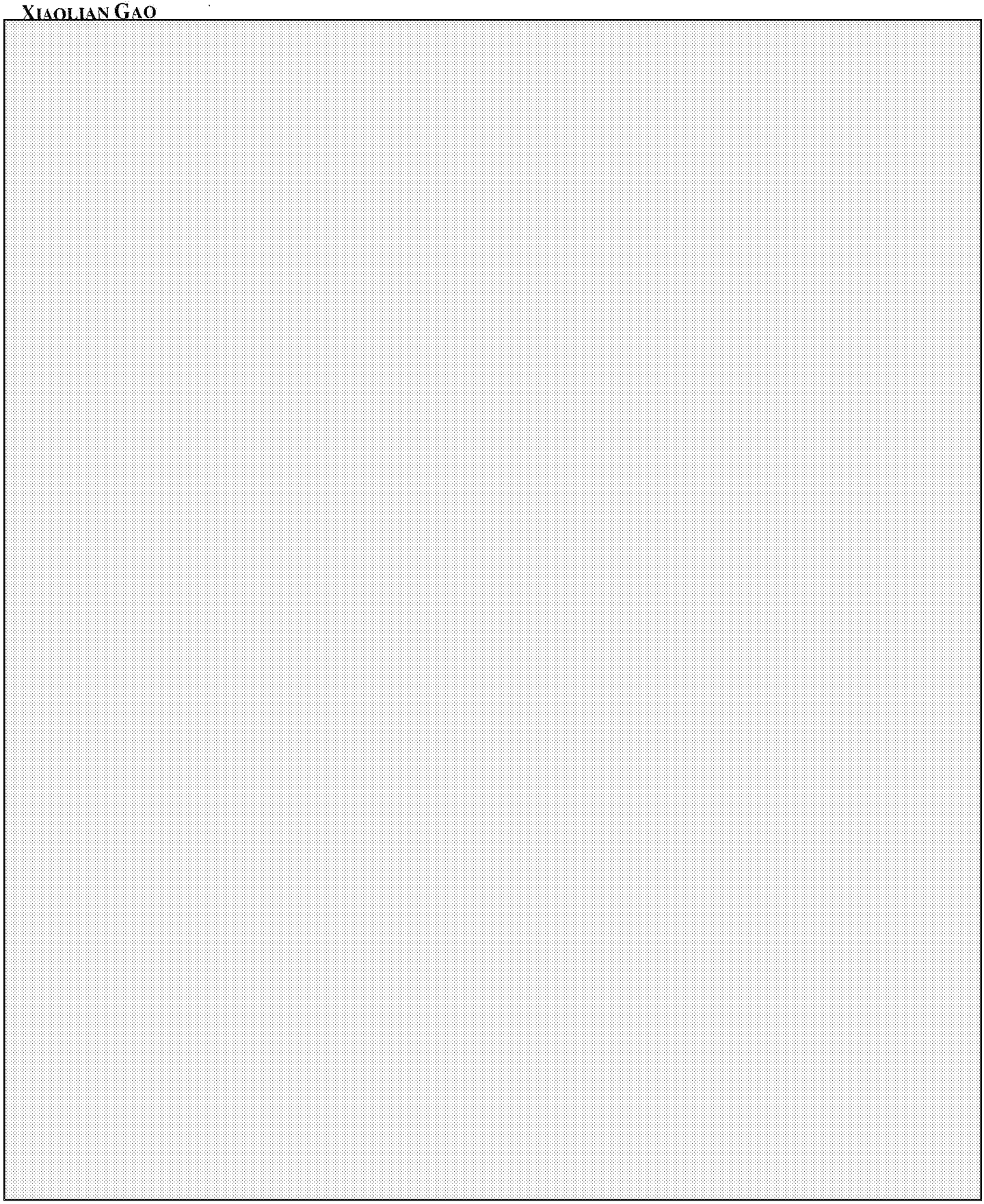


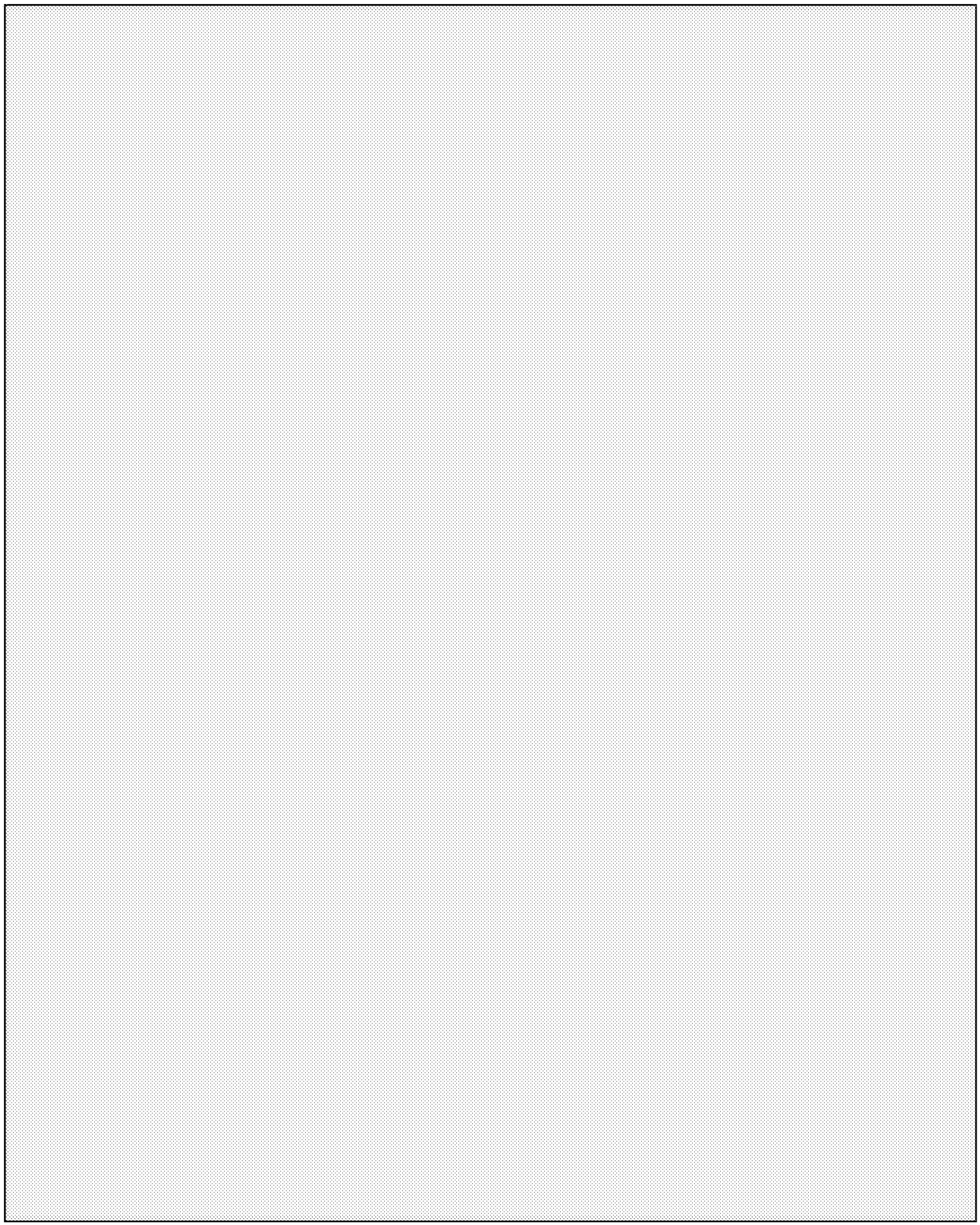


SHIAO-CHUN (DAVID) TU

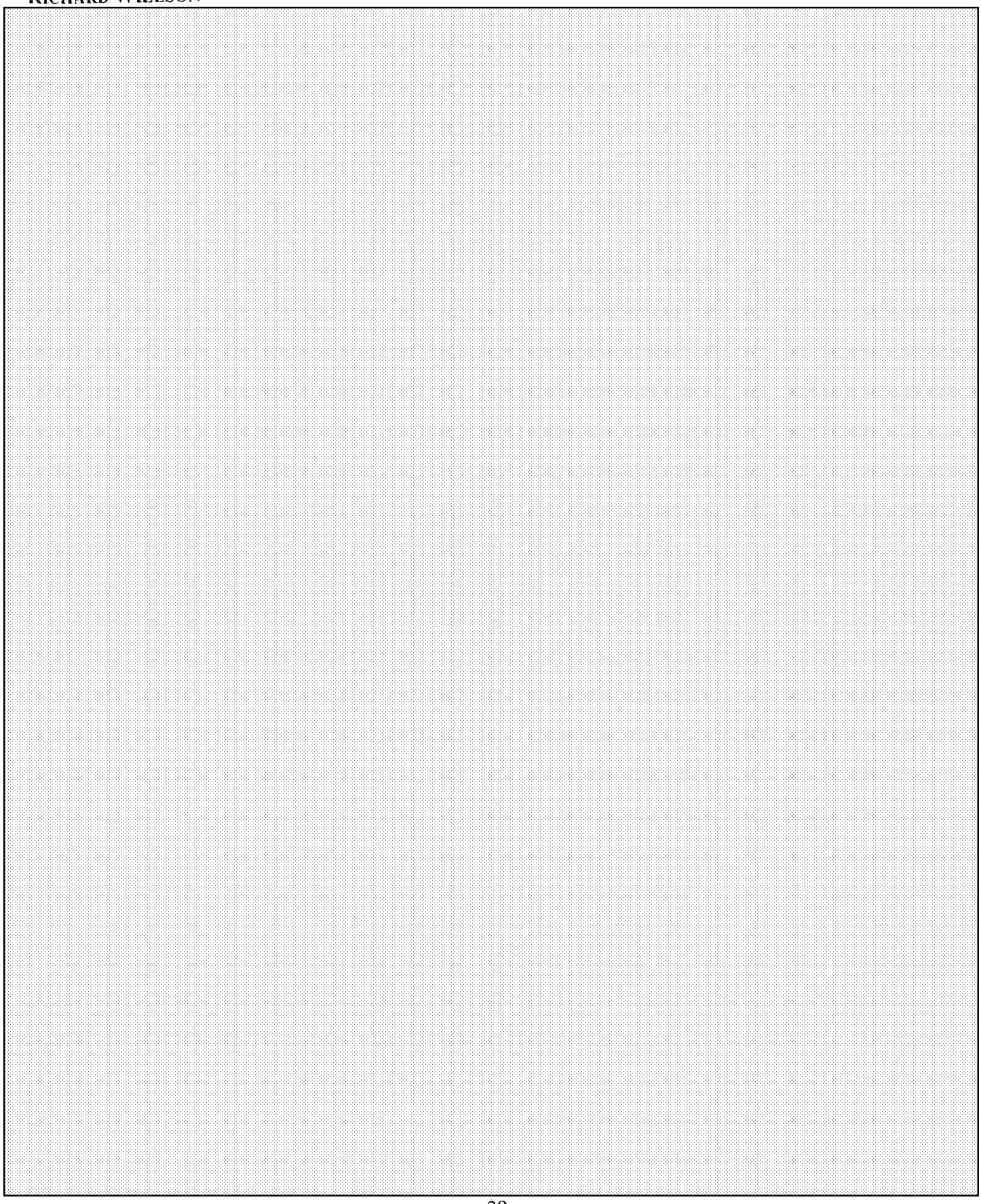


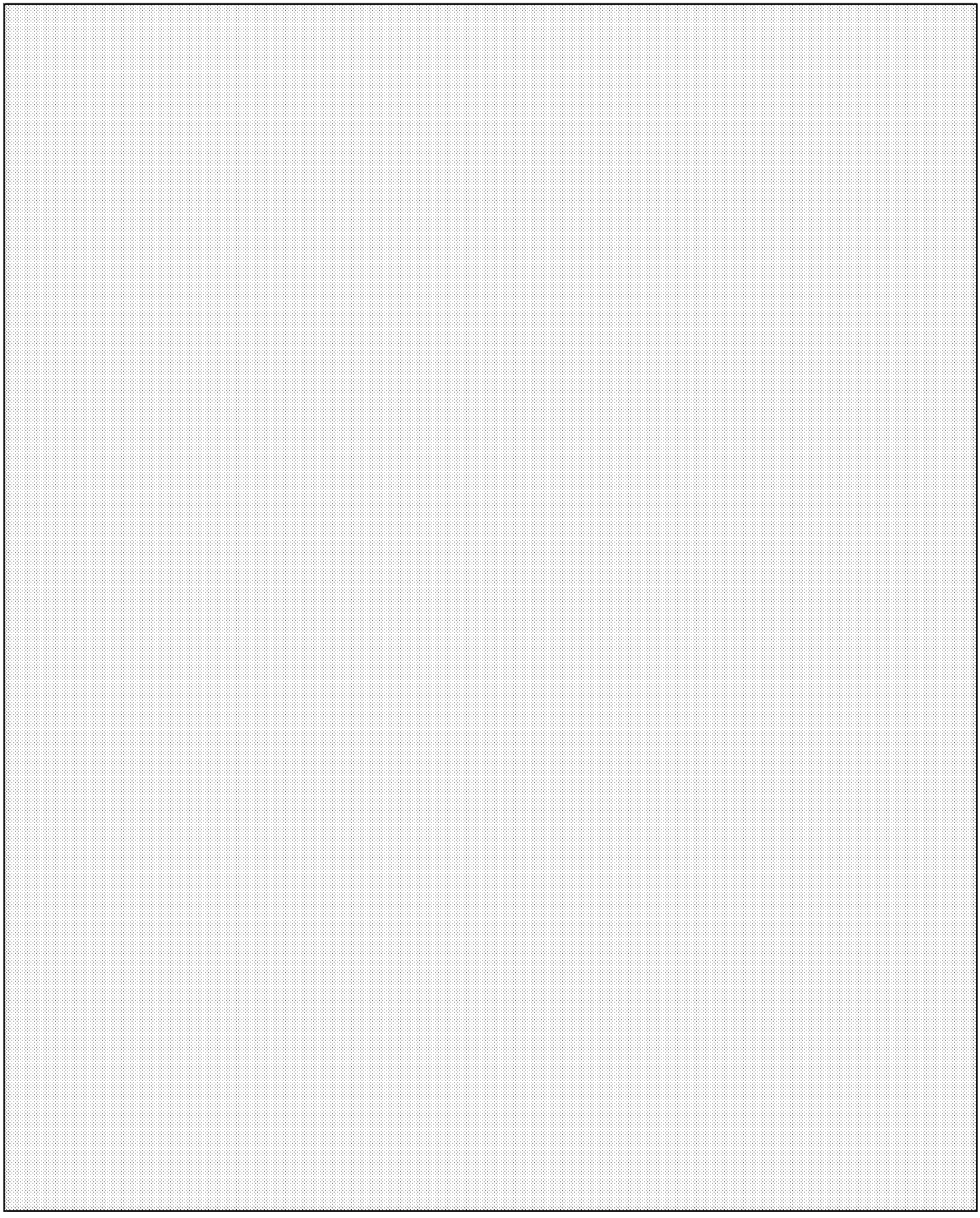




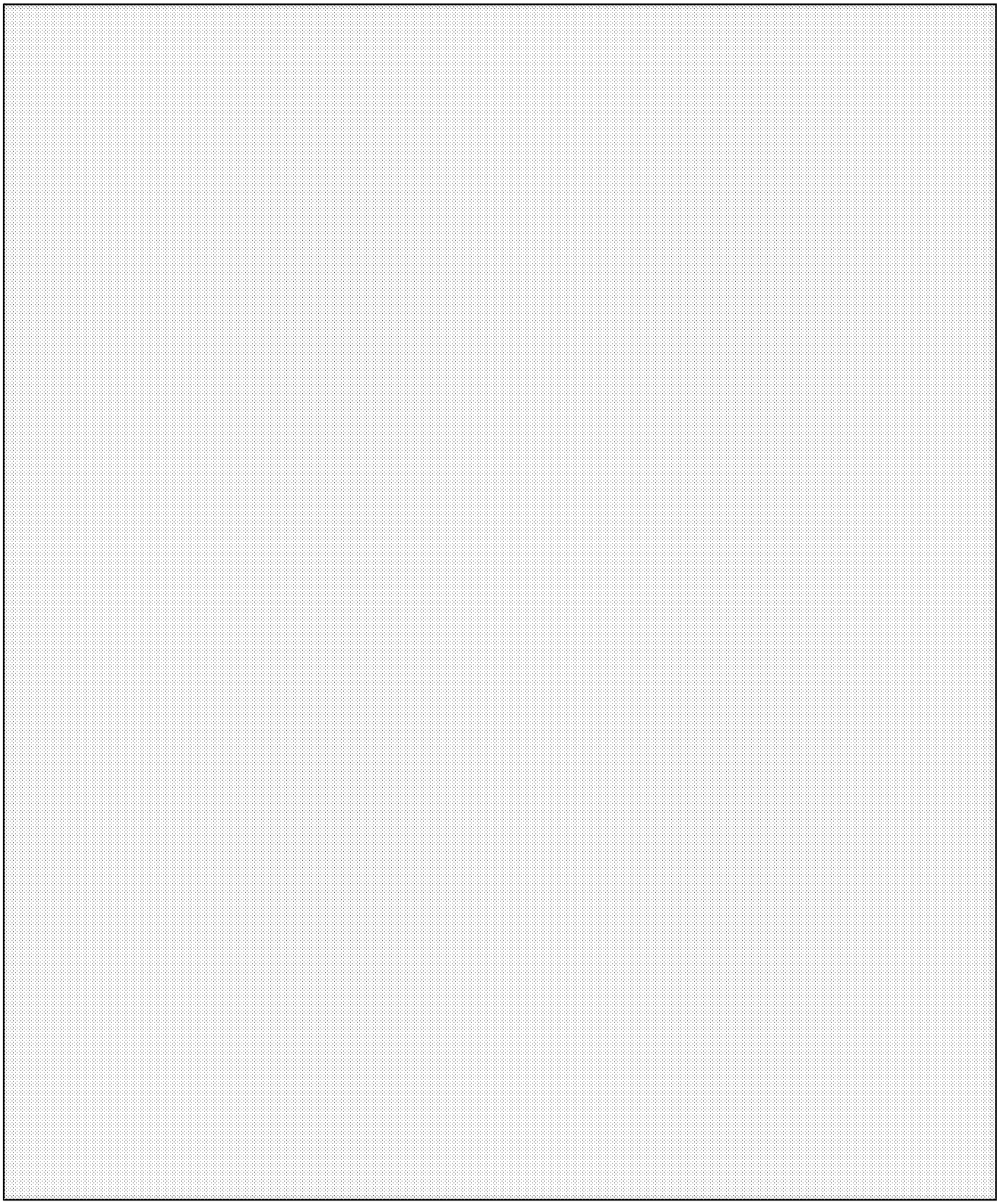


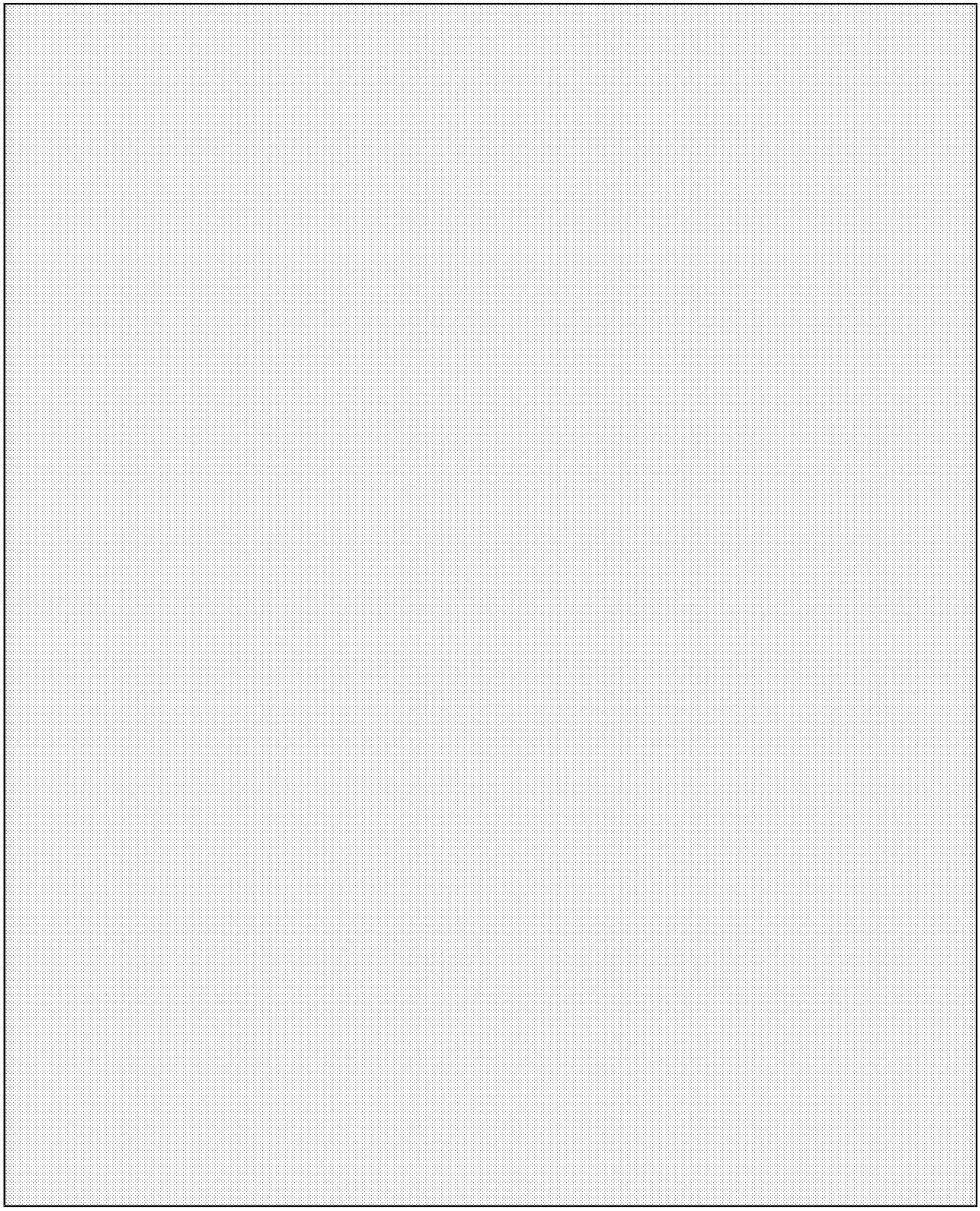
RICHARD WILLSON



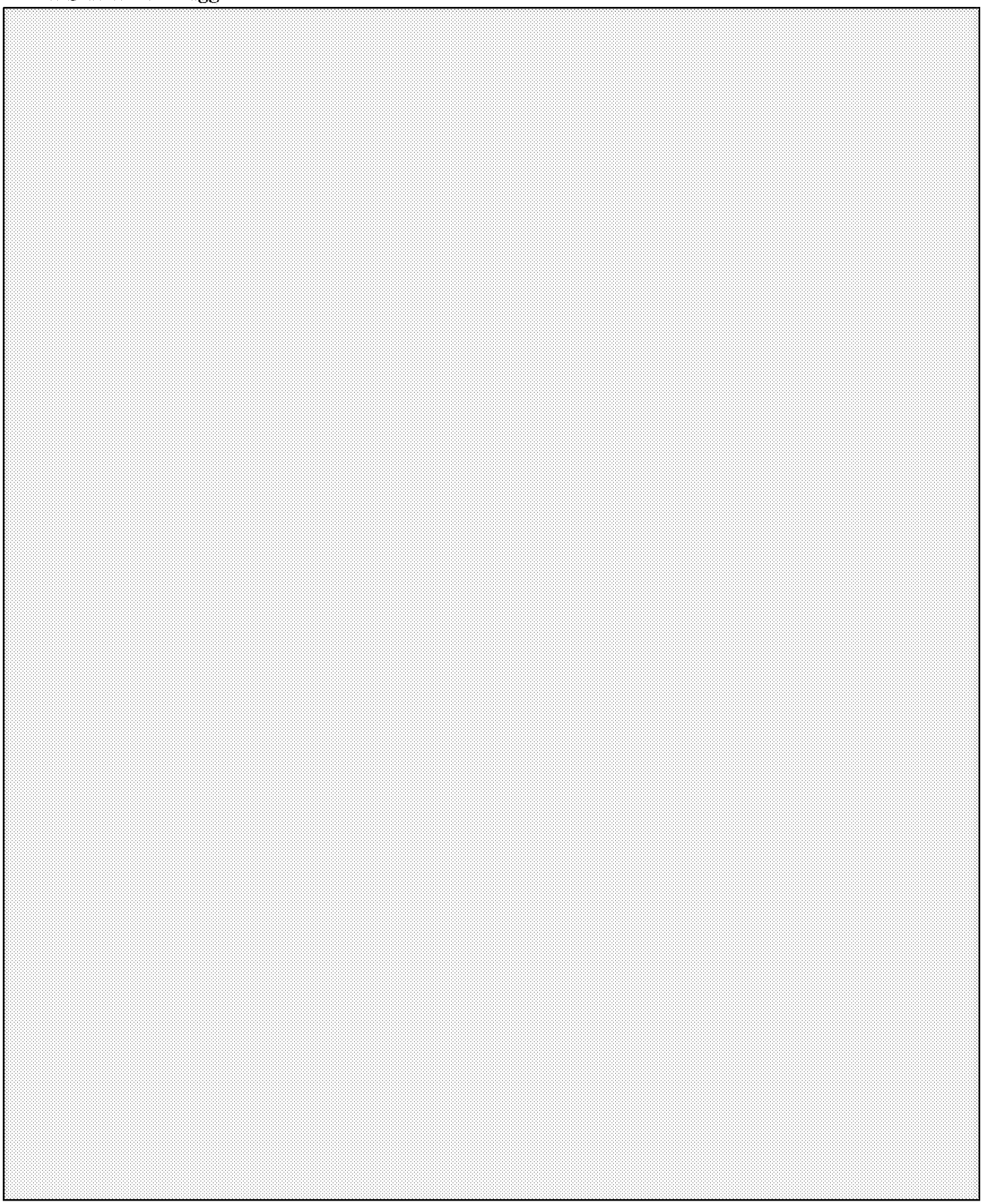


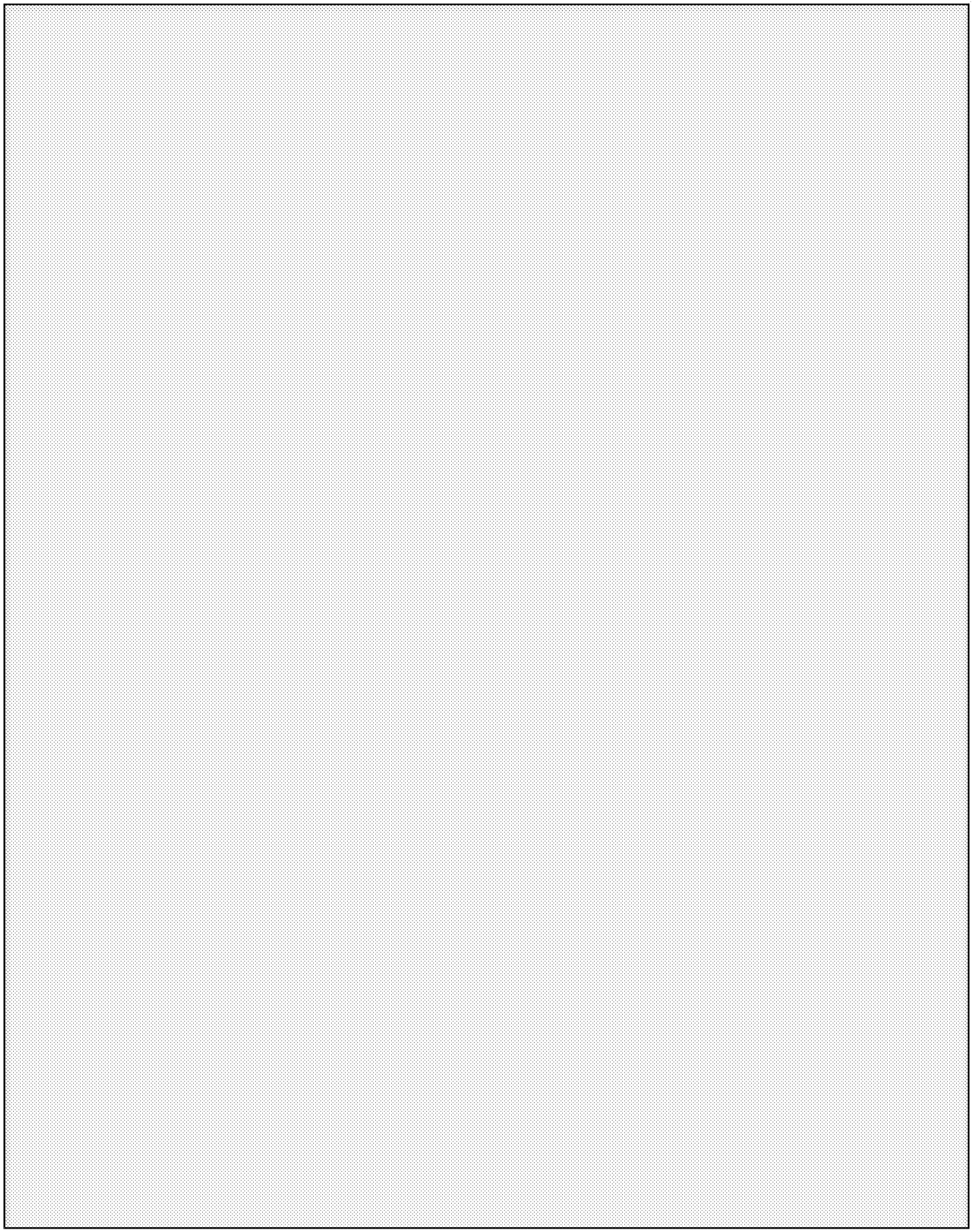
Dr. Susan Hardin



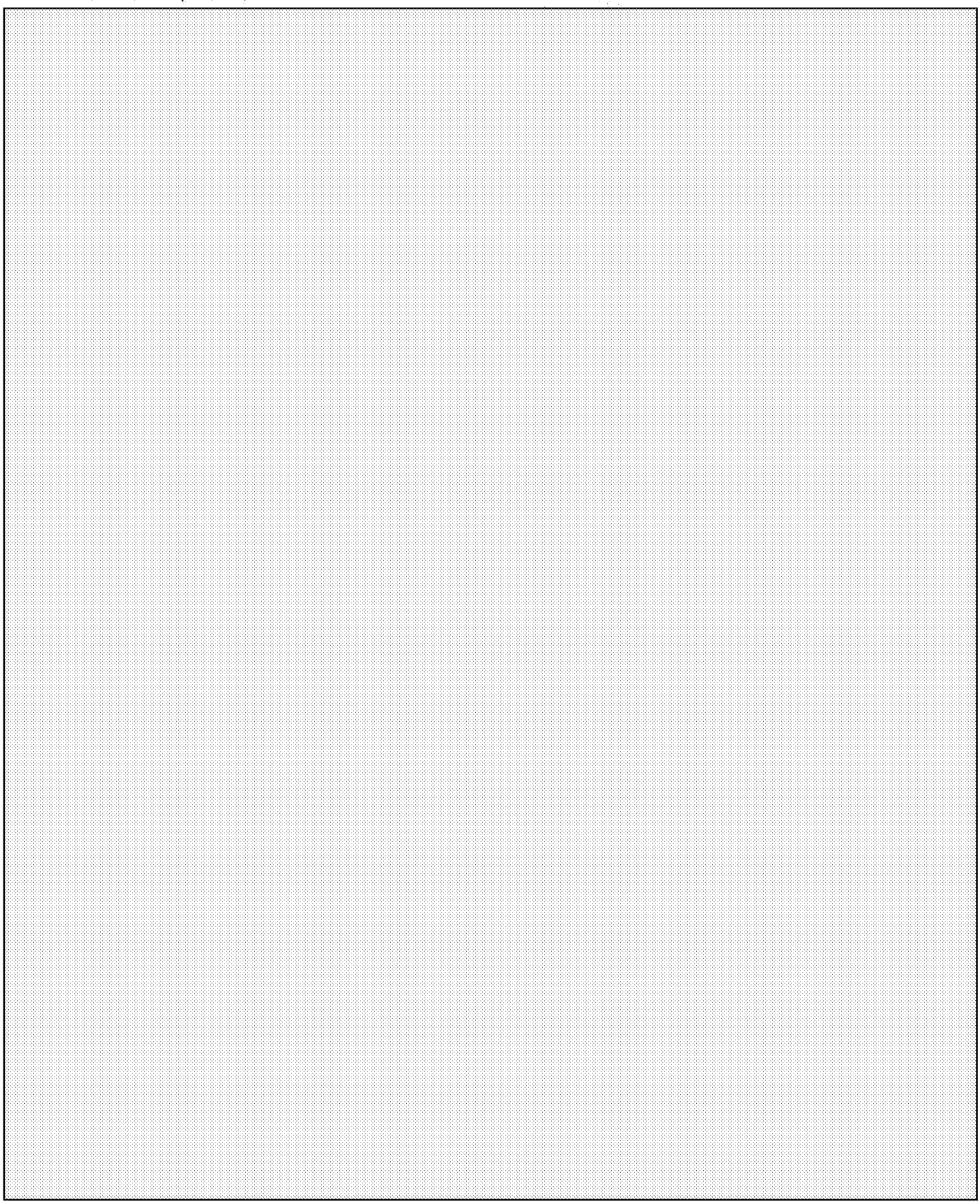


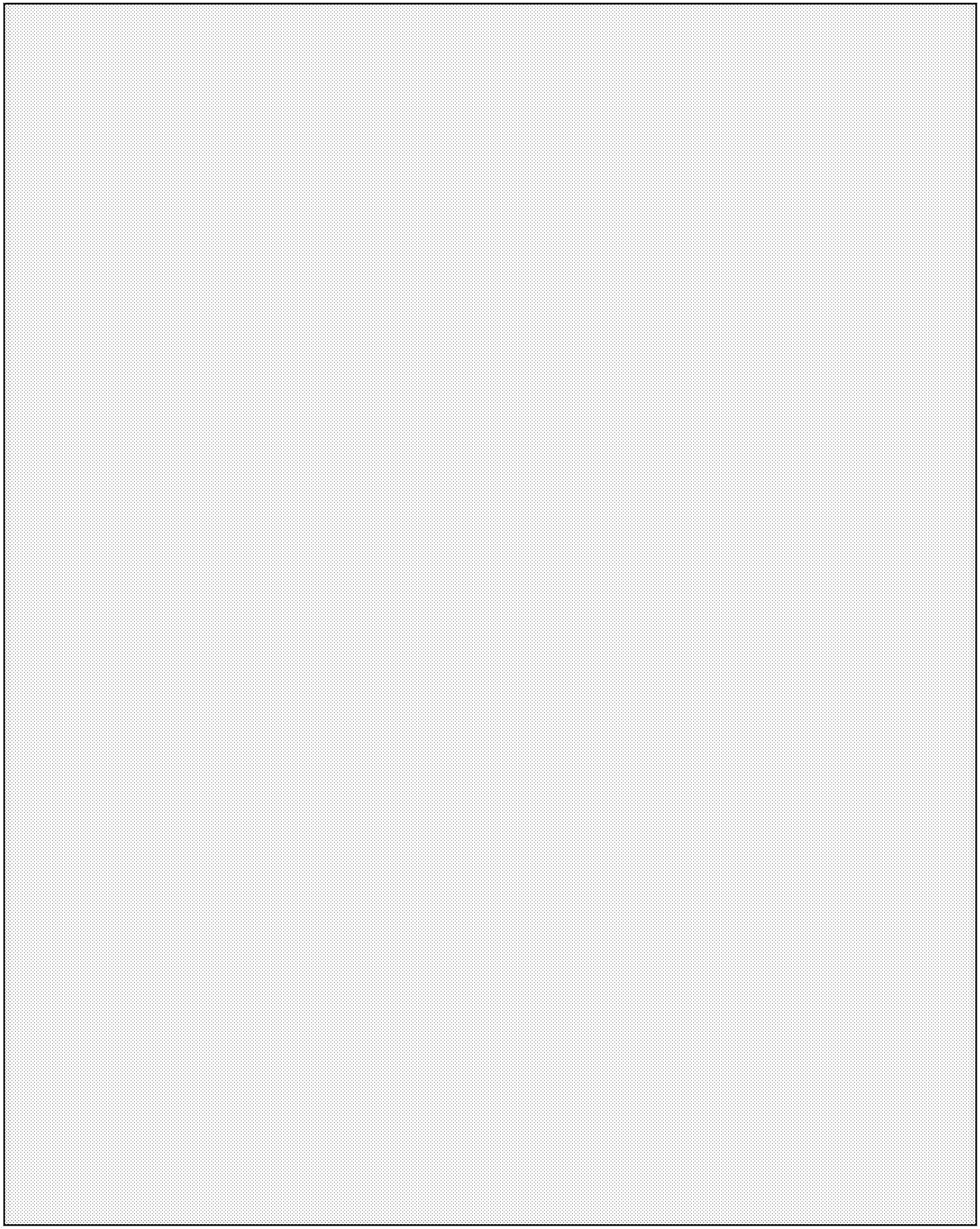
Dr. James M. Briggs



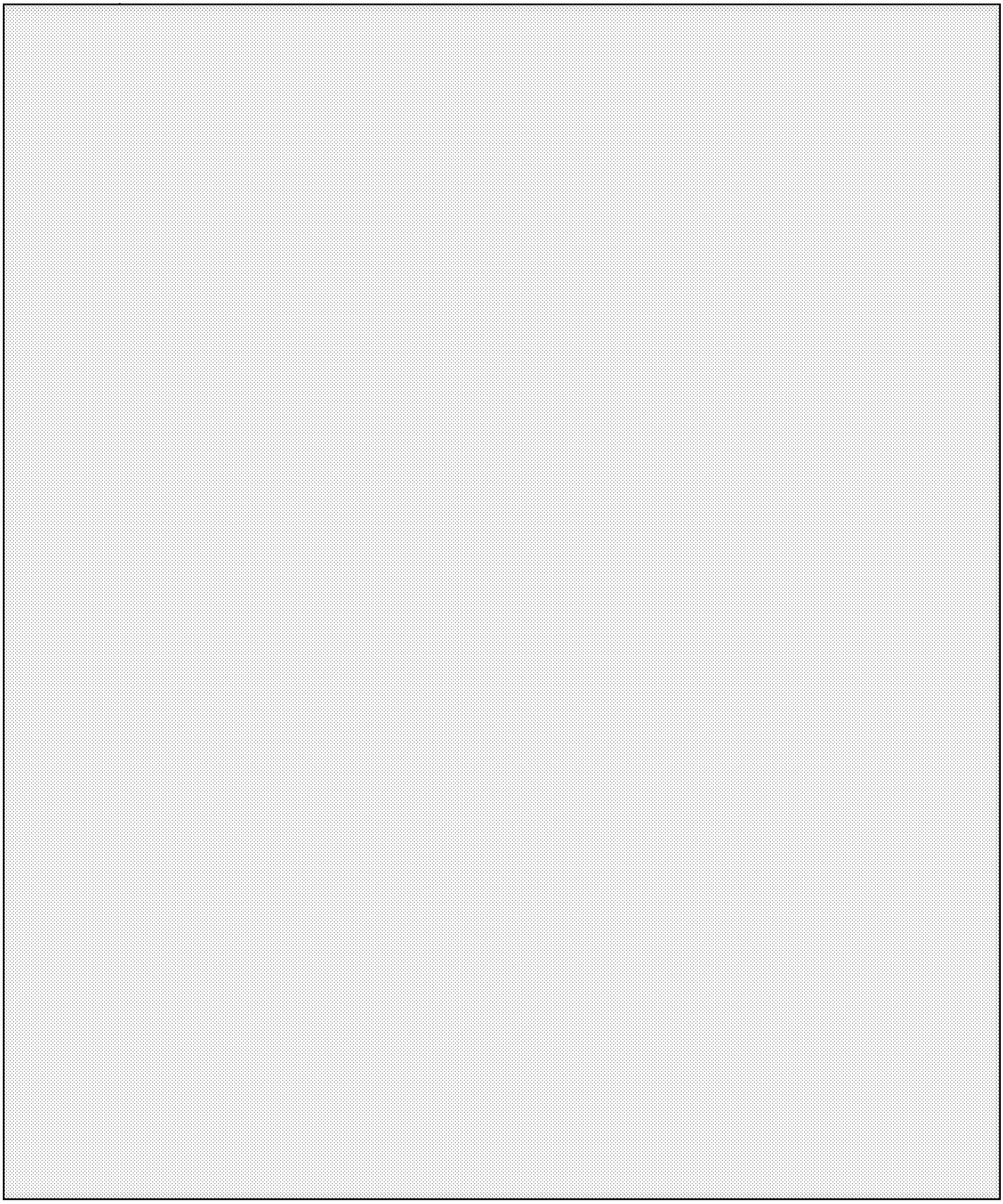


Dr. Shiao-Chun (David) Tu

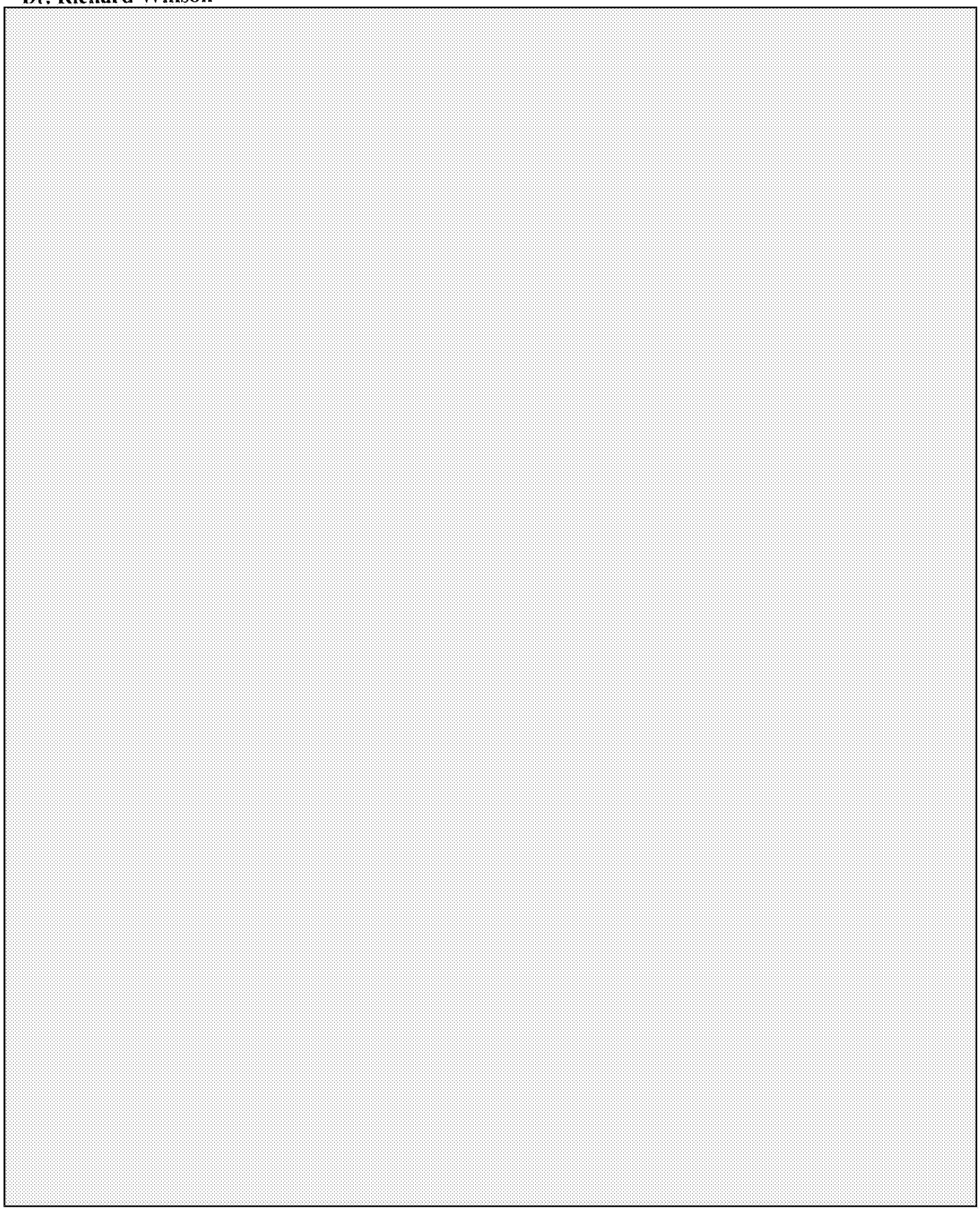


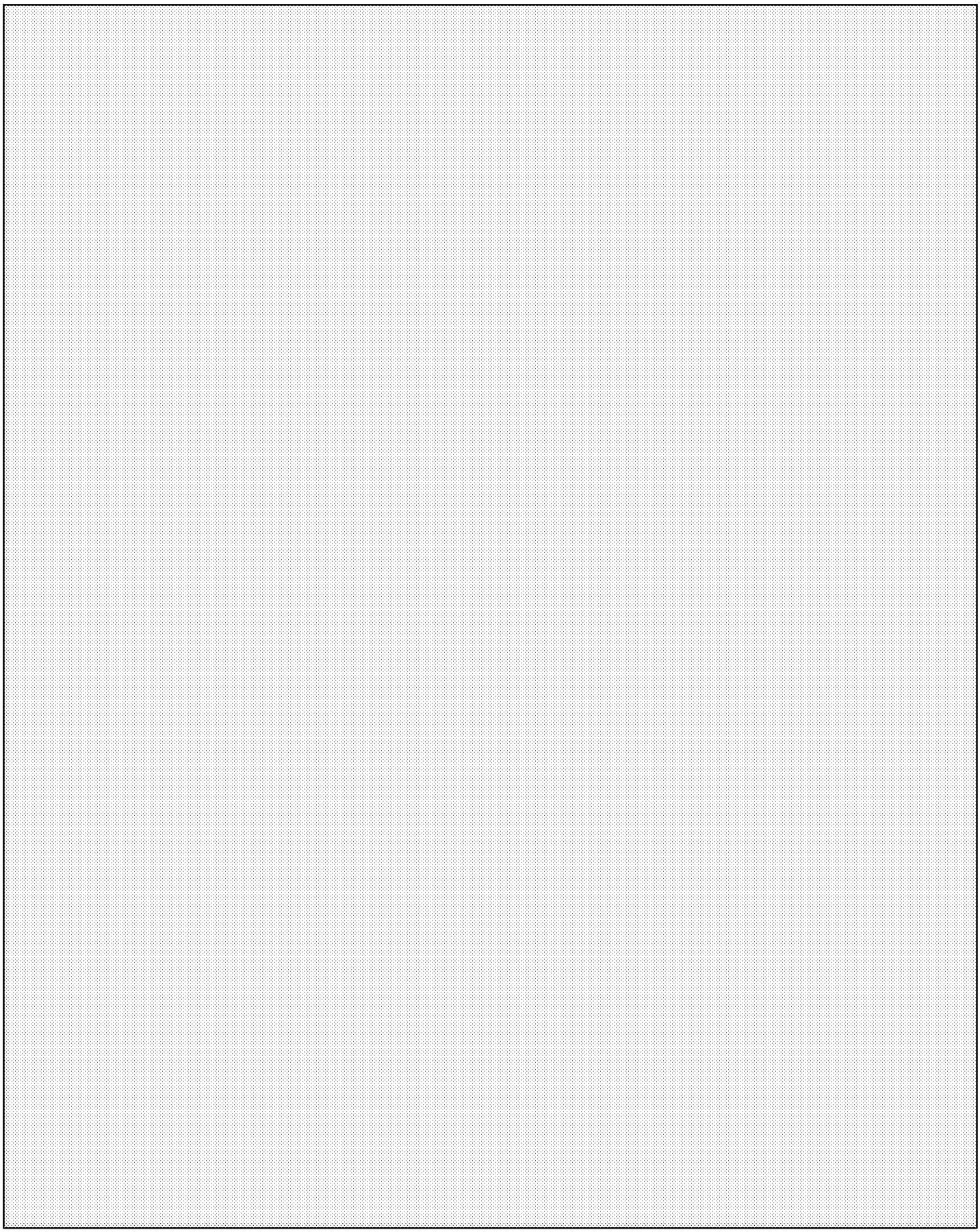


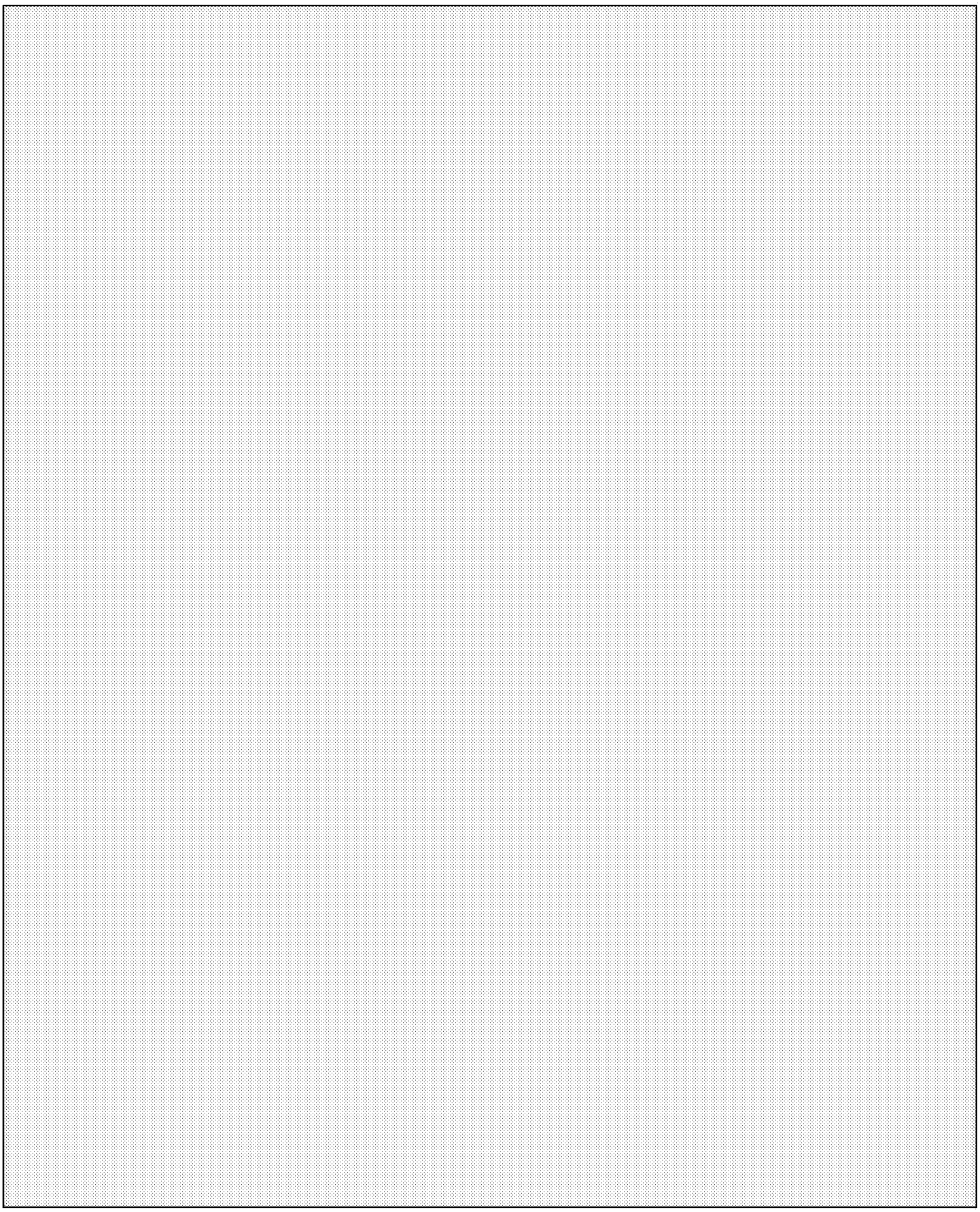
Dr. Xiaolian Gao



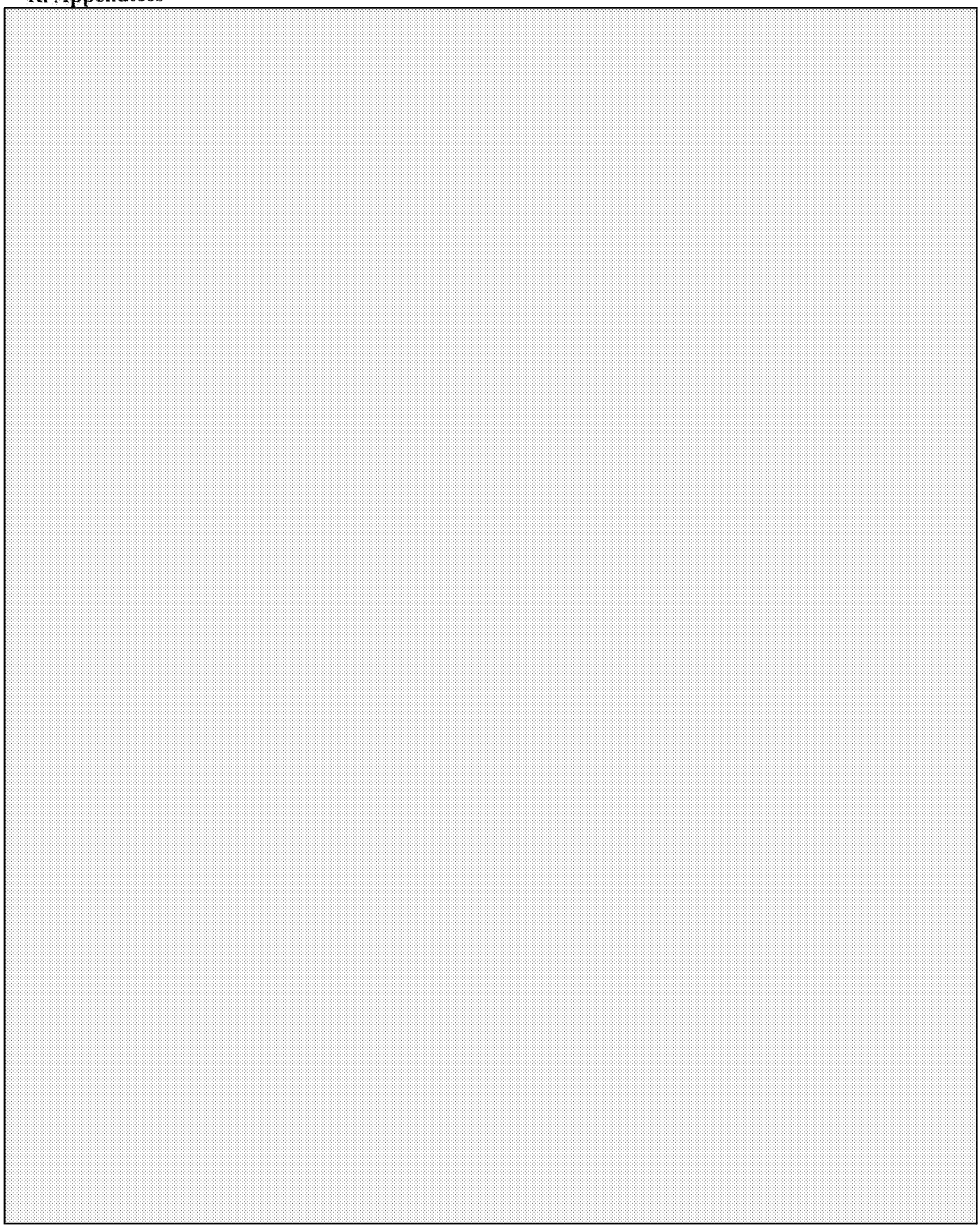
Dr. Richard Willson







K. Appendices



Principal Investigator/Program Director *(Last, first, middle):* Hardin, Susan Houck

DETAILED BUDGET FOR INITIAL BUDGET PERIOD

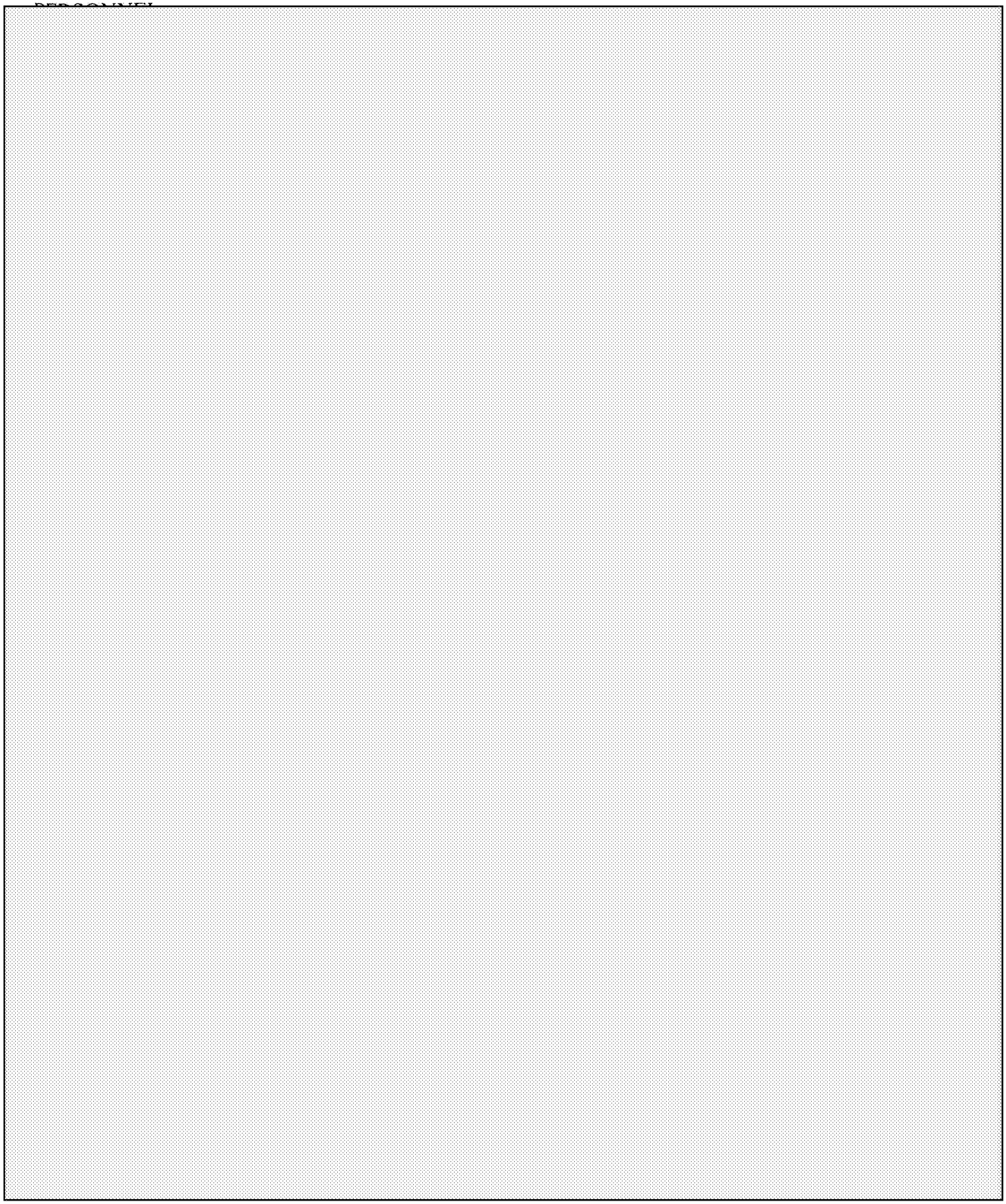
DOLLAR AMOUNT REQUESTED

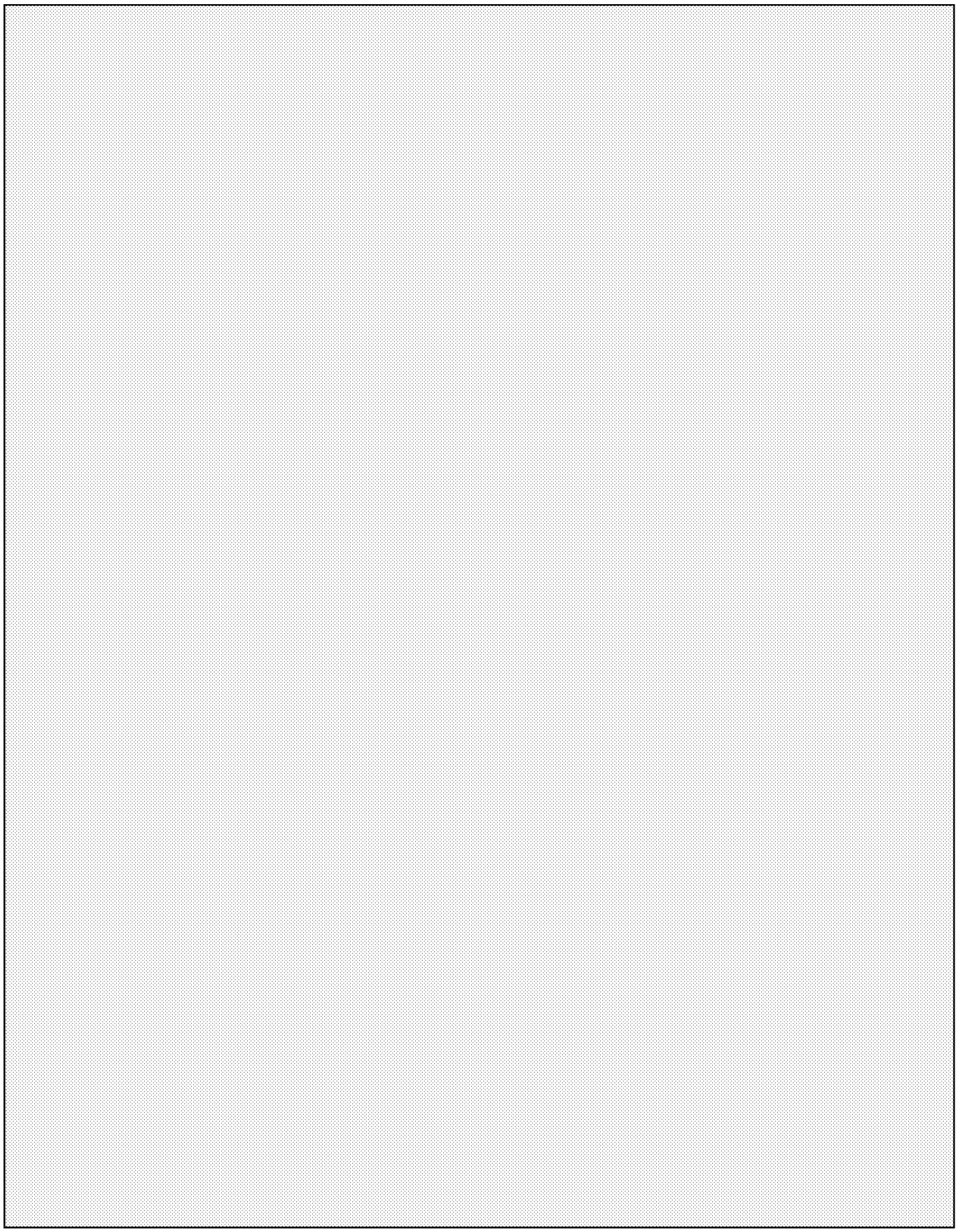
PERSONNEL NAME	ROLE ON PROJECT	TYPE APPT. *(months*	% EFFORT ON PROJ.	INST. BASE SALARY	SALARY REQUESTED	FRINGE BENEFITS	TOTALS

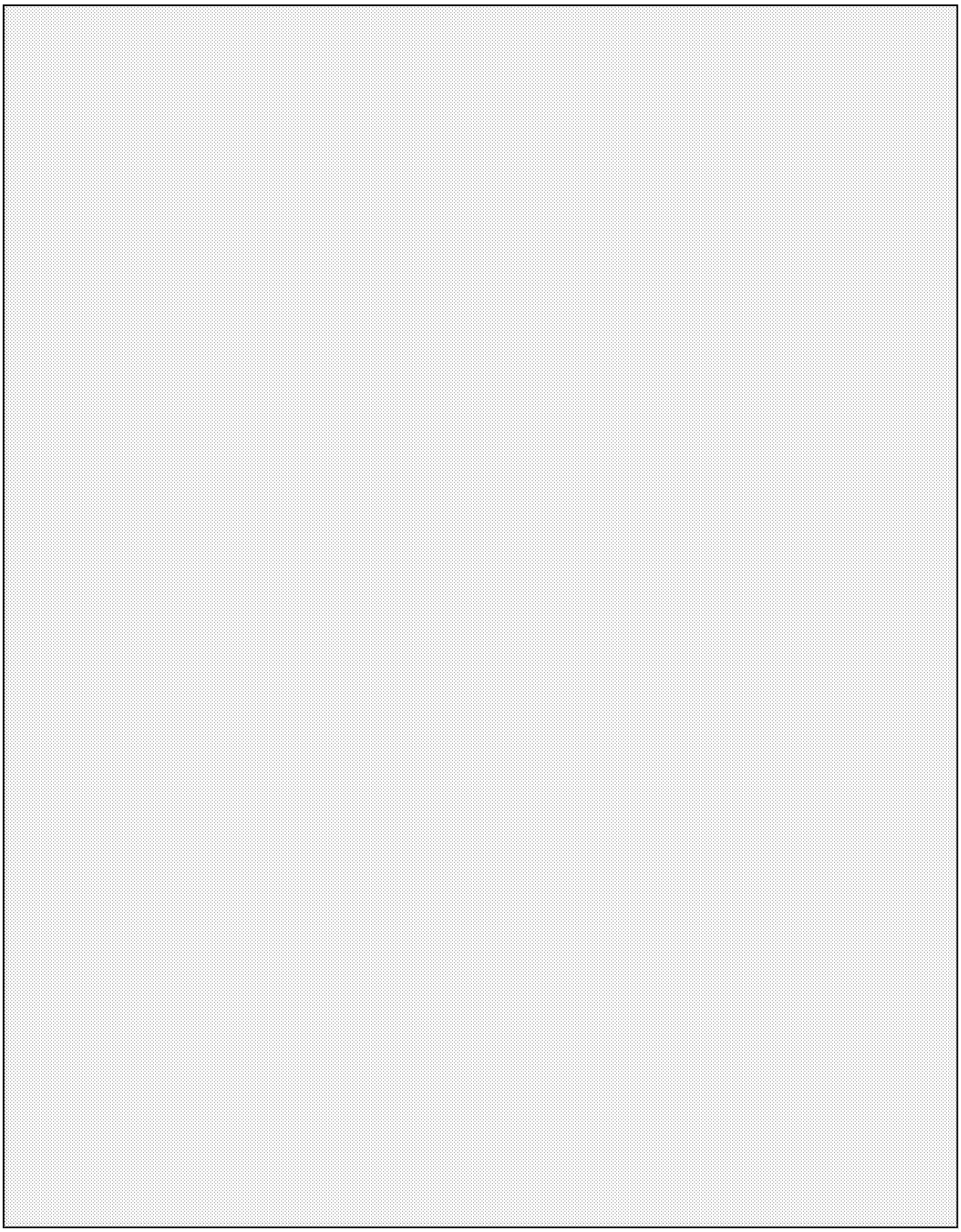
TOTAL COSTS FOR INITIAL BUDGET PERIOD → $845.504

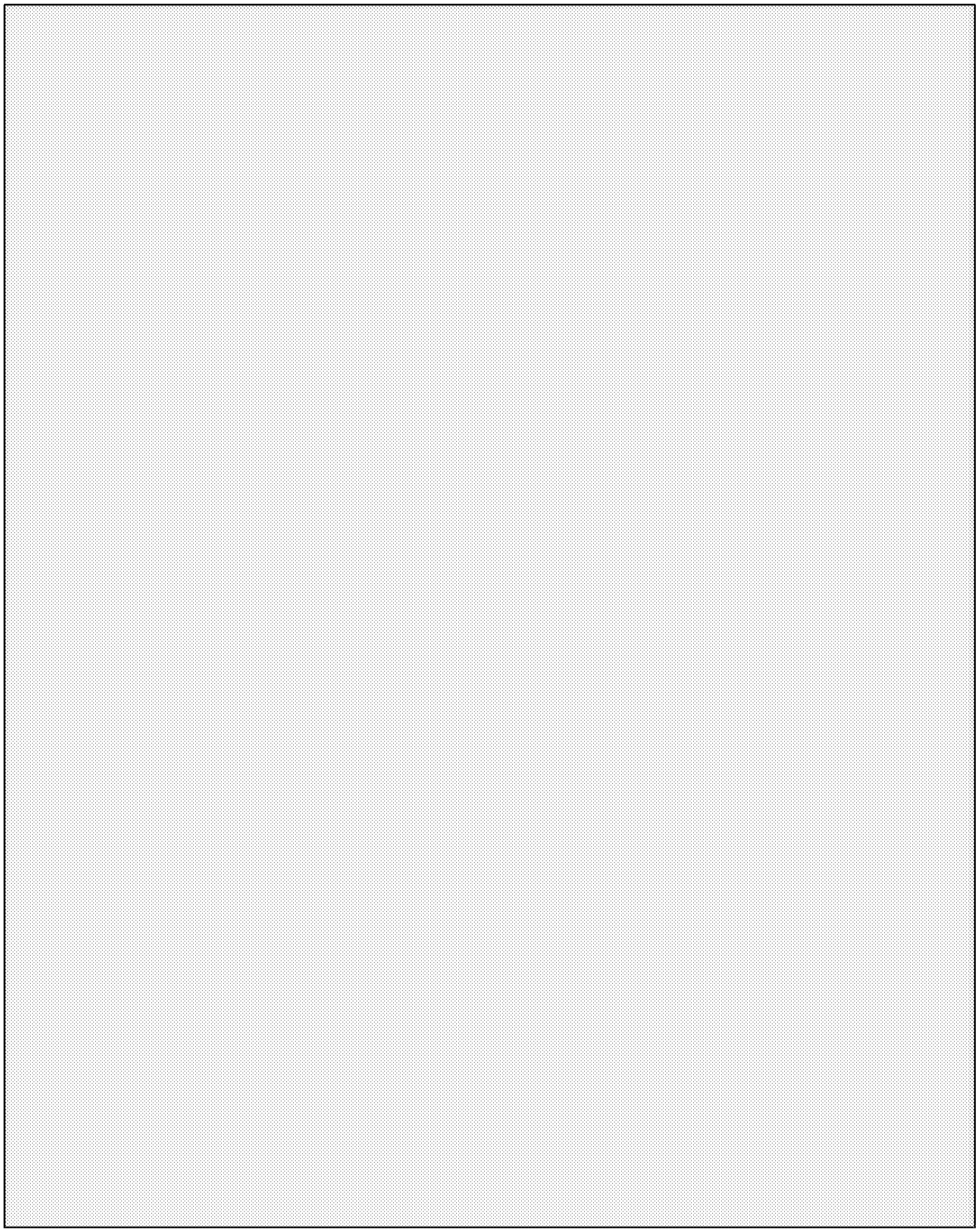
L. Budget Justification

PERSONNEL









UNIVERSITY OF HOUSTON SYSTEM
UNIVERSITY OF HOUSTON

Office of the Vice Chancellor for Research and Intellectual Property Management
UH System

Office of the Vice President for Research
University of Houston